

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 19.4298 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-2
Perfect score: 1485
Sequence: 1 MRGSHHHHHGSMASGRGGL.....YGGLSQGTSGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1485	100.0	291	22	AA882608 Spider recombinant
2	1425	96.0	681	22	AA882609 Spider recombinant
3	1357	91.4	691	22	AA882610 Spider recombinant
4	1037	69.8	989	23	AA50038 N. clavipes spider
5	1036	69.8	606	16	AA99053 Spider dragline va
6	1036	69.8	606	20	AAV40100 Polymer of an anal
7	1031.5	69.5	630	23	AA50042 N. clavipes spider
8	1031.5	69.5	676	23	AA50047 N. clavipes spider
9	1031.5	69.5	1255	23	AA50037 N. clavipes spider
10	1031.5	69.5	1880	23	AA50039 N. clavipes spider

11	1031	69.4	646	18	AAW27178 Nephila clavipes s
12	1015	68.4	651	20	AAV40097 Spider silk protei
13	1015	68.4	651	23	AAU11781 Spider natural sil
14	1015	68.4	718	19	AAV53346 Nephila clavipes s
15	1015	68.4	718	21	AAV59070 N. clavipes spider
16	1008	67.9	528	22	AA82611 Spider recombinant
17	1001	67.4	718	12	AA814308 N.clavipes draglin
18	984	66.3	604	16	AA99057 Spider dragline va
19	977	65.8	364	23	AA50043 N. clavipes spider
20	977	65.8	773	23	AA50045 N. clavipes spider
21	977	65.8	777	23	AA50046 N. clavipes spider
22	976	65.7	606	16	AA99055 Spider dragline va
23	976	65.7	606	20	AAV40101 Polymer of an anal
24	976	65.7	606	20	AAV40102 Polymer of an anal
25	976	65.7	809	23	AAU11793 Dragline protein 1
26	976	65.7	818	23	AAU11797 Dragline protein 1
27	976	65.7	1617	23	AAU11794 Dragline protein 1
28	976	65.7	1626	23	AAU11798 Dragline protein 1
29	922	62.1	271	23	AA50044 N. clavipes spider
30	899	60.5	655	23	AB876672 Protein related to
31	859.5	57.9	230	23	AA50035 N. clavipes spider
32	693	46.7	170	23	AA50041 N. clavipes spider
33	677.5	45.6	615	20	AAV40099 Spider silk protei
34	674	45.4	831	16	AA80168 pMIS1 Misp spider
35	610.5	41.1	641	20	AAV28843 Epstein Barr Virus
36	610.5	41.1	641	21	AAV95856 Epstein Barr virus
37	610.5	41.1	641	22	AA82332 EBV tethering prot
38	593.5	40.0	738	19	AAW56163 NEW DNA sequence 1
39	591.5	39.8	261	19	AAW79137 FLGA Gly-ala inser
40	580	39.1	264	23	AA50048 N. clavipes spider
41	579	39.0	219	23	AA50040 N. clavipes spider
42	561	37.8	235	17	AAW05704 Glycine-rich repea
43	561	37.8	235	19	AAW79126 Epstein Barr Virus
44	554	37.3	531	20	AAV40098 Spider silk protei
45	554	37.3	595	12	AA814309 N.clavipes draglin

ALIGNMENTS

RESULT 1
AA882608
ID AA882608 standard; Protein; 291 AA.
XX AA882608;
AC
XX
DT 02-OCT-2001 (first entry)
XX
DE Spider recombinant silk protein pQE(spl)7.
XX
KW Spider; orb-weaver; silk protein; pQE(spl)7; structural protein;
KW purification; fibre; spinning.
XX
OS Nephila clavipes.
XX
FH Key Location/Qualifiers
FT Misc-difference 232
FT /note= "encoded by GGY"
XX
PN WO200153333-A1.
XX
PD 26-JUL-2001.
XX
PF 01-NOV-2000; 2000WO-US30086.
XX
PR 20-JAN-2000; 2000US-0490291.
XX
PA (MELL/) MELLO C M.
PA (ARCI/) ARCIDIAONO S.
PA (BUTL/) BUTLER M M.
PA (USSA) US SEC OF ARMY.
XX
PI Mello CM, Arcidiacono S, Butler MM;

[illegible]

RESULT 3	
AAB82610	
ID	AAB82610 standard; Protein; 691 AA.
XX	
XX	
AC	
AC	
AAB82610;	
XX	
XX	
DT	02-OCT-2001 (first entry)
XX	
XX	
DE	Spider recombinant silk protein pET((SP1)4/(SP2)1)4.
XX	
XX	
KW	Spider; orb-weaver; silk protein; pET((SP1)4/(SP2)1)4;
KW	structural protein; purification; fibre; spinning.

XX	WO2001I53333-A1.
PN	
XX	
XX	26-JUL-2001.
PD	
XX	
XX	01-NOV-2000; 2000WO-US30086.
PF	
XX	
PR	20-JAN-2000; 2000US-0490291.
XX	
XX	(MELL/) MELLO C M.
PA	(ARCI/) ARCIDIAcono S.
PA	(BUTL/) BUTLER M M.
PA	(USSA) US SEC OF ARMY.
XX	
PI	Mello CM, Arcidiacono S, Butler MM;
XX	
XX	WPI; 2001-483136/52.
DR	N-PSDB; AAH76303.
DR	

Recovering structural polypeptides in a biological sample, useful for purifying and spinning spider silks and other structural proteins, comprises treating the sample containing the polypeptides with an acid.

Claim 2; Page 38-40; 49pp; English.

The present sequence is that of orb-weaver spider (*Nephila clavipes*) recombinant silk protein pET(SPI)4/(SP2)114. The invention provides methods for purifying and spinning spider silks and other structural proteins. Organic acids are used to lyse recombinant cells or other biological samples (such as non-recombinantly derived cells), and enrich the purity and yields of structural proteins by hydrolysing many of the macromolecules while leaving the structural proteins intact. In the case of silk proteins, the resulting lysate is further purified by ion-exchange or affinity chromatography and processed into an aqueous-based mixture for fibre spinning. In the present case, the pET(SPI)4/(SP2)114 gene was cloned into vector pET24 for recombinant expression in *Escherichia coli*, and recombinant silk protein was obtained in 75-75% purity using propionic acid and anion-exchange chromatography on QAE-Sephadex A50. When denaturant (3 M guanidine-HCl) was added to the propionic acid, the recombinant protein was obtained in 80% purity. Products obtained using the methods of the invention can be used in the construction of many materials including films, fibres, woven articles, sutures, ballistic protection, parachutes and parachute cords. The new method has the following advantages over prior art: it involves fewer steps, requires less time and smaller volumes of reagents, results in better recovery of protein at higher purity (70-99%), is easy to scale up, and the fibres are spun in an environmentally benign solution reducing hazardous waste accumulation and cost.

AA
SQ
Sequence 691 AA;

Query Match	91.4%	Score 1357;	DB 22;	Length 691;
Best Local Similarity	95.8%;	Pred. No.	8.8e-98;	
Matches 272;	Conservative 0;	Mismatches 0;	Indels 12;	Gaps 1;
Qy	11	GSMASGRGGLGGQGAGAAAAA	AAAAGAGAGGGYGGGLGSQGTSGRGGLGGGAGGAGAAAAA	70
Dd	13	GSMASGRGGLGGQGAGAAAAA	AAAAGAGAGGGYGGGLGSQGTSGRGGLGGGAGGAGAAAAA	72
Qy	71	AAAAGGACGGCGYGGGLGSQGTSGRGGLGGAGAAAAA	AAAAGAGAGCGGYGGLGSQGTSG	130
Dd	73	AAAAGGACGGCGYGGGLGSQGTSGRGGLGGAGAAAAA	AAAAGAGAGCGGYGGLGSQGTSG	132
Qy	131	RGGLGGQGAGAAAAA	AAAAGAGAGCGGYGGGLGSOG-----TSRGGGLGGQGAG	178
Dd	133	RGGLGGQGAGAAAAA	AAAAGAGAGCGGYGGGLGSQGTSGPGCYGPCCQTSGRGGLCGGCAG	192
Qy	179	AAAAAAAAAGGAGCGGYGGGLGSQGTSGRGGLGGGAGAAAAA	AAAAGAGAGCGGYGGGLG	238
Dd	193	AAAAAAAAAGGAGCGGYGGGLGSQGTSGRGGLGGGAGAAAAA	AAAAGAGAGCGGYGGGLG	252
Qy	239	SQGTSGRGGLGGQGAGAAAAA	AAAAGAGAGCGGYGGGLGSQGTSG	282
Dd	253	SQGTSGRGGLGGQGAGAAAAA	AAAAGAGAGCGGYGGGLGSQGTSG	296

RESULT 4	
AAM50038	
ID	AAM50038 standard; protein; 989 AA.
XX	
XX	
AC	AAM50038;
XX	
XX	
DT	18-SEP-2002 (first entry)
XX	
DE	N. clavipes spidroin synthetic homologue SolSM12 protein.
XX	
XX	
KW	Spidroin; spider; silk; fibre; film; membrane; wound; fill.
XX	
XX	
OS	Synthetic.
XX	
XX	
FN	DE10113781-A1.
XX	
XX	
PD	13-DEC-2001.

21-MAR-2001; 2001DE-1013781.

XX
XX

PR 09-JUN-2000; 2000DE-1028212.
PR 24-OCT-2000; 2000DE-1053478.

XX
PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.XX
PI Scheller J, Conrad U, Grosse F, Guehrs K;

XX
DR WPI: 2002-123561/17.

DR N-PsDB; ABL61039.

XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
PT wounds, comprises modules that encode repeating units of spider
PT proteins -
PT proteins -

PS Claim 22; Page 34-38; 88pp; German.

This invention describes a novel DNA sequence, encoding a synthetic spider silk protein, comprising modules, each comprising a group of sequentially arranged oligonucleotides, each oligonucleotide encoding a repeating unit of a spider protein. The synthetic protein has at least 84% homology with the Nephila clavipes spider protein and is used to produce synthetic fibres, films and/or membranes, particularly: (i) for medical use, especially to close wounds and/or to support or cover artificial organs; (ii) as adhesion surfaces for culturing cells; and (iii) as fillers. The synthetic proteins are very similar to native spider silk proteins; can be prepared on a large scale and can be spun to fibres with excellent mechanical properties (strength and elasticity).

CC Also they retain water solubility after long-term boiling in aqueous
CC solutions and since they are also soluble in organic solvents but
CC precipitated at high salt concentration, they are easily extracted and
CC purified. The modular construction of the invention facilitates
CC incorporation of additional peptide-encoding sequences, e.g. to simplify
CC purification or modulate solubility. This sequence represents the
CC synthetic N. clavipes spidroin-1 homologue SolSM12 described in the
CC invention.
XX
SQ Sequence 989 AA;

Query Match 69.8%; Score 1037; DB 23; Length 989;
Best Local Similarity 50.7%; Pred. No. 7e-73; Mismatches 7; Indels 228; Gaps 13;
Matches 250; Conservative 7;

QY 11 GSMASGRGGLGGCGAGAAA-----AAAAAGGAGGGY 44
|| :|||||
Db 345 GSQAGRGGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGAGAAAAGGAGGGY 404

QY 45 GGLGSGTSGRGLGGCGAGAAA-----AAAAAGGAG 78
||||| :|||||
Db 405 GGLGSGTSGRGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGAGAAAAGGAG 463

QY 79 QGGYGGGLGSGT-----SGRGL 96
|||||
Db 464 QGGYGGGLGSGTSGRGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGL 523

QY 97 GGGAGAAA-----AAAAAGGAGGGYGGGLGSGAGRGGL 128
||||| :|||||
Db 524 GGGAGAAAAGGAGGGYGGGLGSGAGRGGLGSGAGRGGGYGGGLGSGAGRGGL 593

QY 129 -----SGRGLGGCGAGAAA-----AGGAGGGY 158
||||| :|||||
Db 584 AAAAAAGGAGGGYGGGLGSGAGRGGLGSGAGRGGGYGGGLGSGAGRGGGY 643

QY 159 GGLGSGTSGRGLGGCGAGAAA-----AAAAAGGAGGGY 198
||||| :|||||
Db 644 GGLGSGTSGRGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGYGGGLGSGAGRGGGY 702

QY 199 LGSQGTSGRGLGGCGAGAAA-----AAAAAGGAGGGY 232
||||| :|||||
Db 703 LGSQGTSGRGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGYGGGLGSGAGRGGGY 761

QY 233 GYGLGSGTSGRGLGGCGAGAAA-----AAAAAGG 266
||||| :|||||
Db 762 GYGLGSGTSGRGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGYGGGLGSGAGRGGGY 820

QY 267 AGGGYGGGLGSG 279
|||||
Db 821 AGGGYGGGLGSG 833

RESULT 5
AAR99053
ID AAR99053 standard; Protein; 606 AA.
XX
AC AAR99053;
XX
DT 17-JAN-1997 (first entry)
XX
DE Spider dragline variant, DP-1A.9 polymer.
XX
KW Spider; dragline protein; variant; monomer; polymer;
KW fibre forming region; Spidroin 1; Nephila clavipes; DP1; mimic;
KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KW rope; surgical suture; implant; reinforcement; film; coating.
XX
OS Synthetic.
XX
PN W09429450-A2.
XX
PD 22-DEC-1994.
XX

PF 15-JUN-1994; 94WO-US06689.
XX
PR 15-JUN-1993; 93US-0077600.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Fahnestock SR;
XX
DR WPI; 1995-036479/05.
XX
XX New synthetic variants of spider dragline protein - for making
PT fibres useful as clothing, surgical silk, plastic reinforcement
PT etc., also related DNA, vectors and transformed cells
XX
PS Disclosure; Page 85-87; 168pp; English.
XX
CC This sequence represents a synthetic spider dragline variant polymer,
CC DP-1A.9. The sequence of the DP-1A.9 monomer is given in AAR99052.
CC The polypeptide monomer is a variant based on a consensus sequence
CC derived from the fibre forming regions of spider dragline protein,
CC esp. the natural protein 1 (Spidroin 1) from Nephila clavipes.
CC DNA sequence encoding the monomer may be used in the recombinant
CC production of the variant protein in a recombinant host, e.g. E. coli
CC or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic
CC the repeating consensus sequence of the natural protein and the pattern
CC of variation among individual repeats. DP-1A analogues are composed
CC of a tandemly repeated 101 amino acid monomer which comprises four
CC repeats which differ from the consensus sequence given in AAR99052,
CC according to the pattern (1)-(5) given below. This 101 amino acid
CC monomer is repeated 1-16 times in a series of analogue proteins. The
CC individual repeats differ from the consensus according to the pattern:
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGRGGLGGCGAGGG;
CC encompasses integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.
XX
SQ Sequence 606 AA;

Query Match 69.8%; Score 1036; DB 16; Length 606;
Best Local Similarity 64.9%; Pred. No. 5.6e-73;
Matches 239; Conservative 4; Mismatches 23; Indels 102; Gaps 13;

QY 11 GSMASGRGGLGGCGAGAAA-----AAAAAGGAGGGYGGGLG 49
|| :|||||
Db 230 GSQAGRGGLGGCGAGAAAAGGAGGGYGGGLGSGAGRGGGYGGGLGSG 289

QY 50 QGT--SGRGLGGCGA---GAAAAAAGGAGGGYGGGLGSGTSGRGLGGCGAGAA 104
|| :|||||
Db 290 QGAGGGYGGGLGSGAGRGGGYGGGLGSGTSGRGLGGCGAGAA 348

QY 105 AA-----AAAAAGGAGGGYGGGLGSGT--SGRGLGGCGA--- 139
|| :|||||
Db 349 AAAAAAGGAGGGYGGGLGSGAGRGGGYGGGLGSGAGGGYGGGLGSGAGR 408

QY 140 -GAAAAAAGGAGGGYGGGLGSGTSGRGLGGCGAGAAA----- 182
|| :|||||
Db 409 GGGAGAAAAGGAGGGYGGGLGSGTSGRGLGGCGAGAA 467

QY 183 -----AAAAAGGAGGGYGGGLGSGT--SGRGLGGCGA---GAAAAAAGGAGGGY 232
|| :|||||
Db 468 GGGAGAAAAGGAGGGYGGGLGSGAGGGYGGGLGSGAGRGGGYGGGLGSGAGGGY 527

QY 233 GYGLGSGTSGRGLGGCGAGAAA-----AAAAAGGAGGGY 271
||||| :|||||

```

Db 528 GYGGLGSQG-AGRGGLGGGAGAAAAAGGAGGGGGLGSQAGGAGAGAAAAAGGAGGG 586
QY 272 YGGLGSQG 279
Db 587 YGGLGSQG 594

RESULT 6
AAY40100
ID AAY40100 standard; protein; 606 AA.
XX
XX AAY40100;
XX
XX 19-NOV-1999 (first entry)
XX
XX Polymer of an analogue of spider silk protein spidroine major 1.
XX
XX Spider silk protein; spidroine major 1; cosmetic; make-up;
XX dermatological compositions; hair care; skin care; sunscreen;
XX hormone; moisturizer; skin disorder; skin disorder.
XX
XX Synthetic.
XX Nephila clavipes.
XX
XX Key Location/Qualifiers
XX Peptide 1..101
XX /note= "monomer unit"
XX
XX FR2774588-A1.
XX
XX 13-AUG-1999.
XX
XX 11-FEB-1998; 98FR-0001614.
XX
XX 11-FEB-1998; 98FR-0001614.
XX
XX (OREA ) L'OREAL SA.
XX
XX Philippe M, Garson JC, Arraudeau JP;
XX WPI; 1999-510729/43.
XX
XX Cosmetic or dermatological composition containing spider silk protein,
XX for hair or skin care, in make-up or sunscreens
XX
XX Claim 8; Flg 4B; 32pp; French.
XX
XX The present sequence represents a polymer of an analogue of the spider
XX silk protein spidroine major 1. The protein improves the moisturizing/
XX softening action of the compositions. The protein, and its fragments
XX are used in cosmetic or dermatological compositions. These compositions
XX have use as hair or skin care products; and make-up or sunscreens.
XX As the protein is a good, persistent film-formers on the skin
XX of low surface density, it can be used for delivery of active
XX agents that are generally difficult to administer, e.g. vitamins,
XX hormones, moisturizers or agents for treating disorders of the
XX skin and hair.
XX
XX Sequence 606 AA;
XX
XX Query Match 69.8%; Score 1036; DB 20; Length 606;
XX Best Local Similarity 64.9%; Pred. No. 5.6e-73;
XX Matches 239; Conservative 4; Mismatches 23; Indels 102; Gaps 13;

QY 11 GSMASRGGLGGGAGAGAAAA-----AAAAAGGAGGGYGGGLGS 49
Db 230 GSGCAGRGGLGGGAGAGAAAAAGGAGGGGGLGSQAGGAGAAAAAGGAGGGYGGGLGS 289
QY 50 QGT--SGRGLGGGGA---GAAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAA 104
Db 290 QGAGGGYGGGLGSQAGRGGGAGAAAAAGGAGGGYGGGLGSQ-AGRGGLGGGAGAA 348
QY 105 AA-----AAAAAGGAGGGYGGGLGSQGT--SGRGLGGGGA-- 139

```

```

Db 349 AAAAAGAGGGGLGSQAGGAGAAAAAGGAGGGYGGGLGSQAGGAGGCGGGLGSQAGR 408
QY 140 -GAAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAGAAA----- 182
Db 409 GCGCAGAAAAAGGAGGGYGGGLGSQ-AGRGGLGGGAGAAAAAGGAGGGYGGGLGSOGA 467
QY 183 -----AAAAAGGAGGGYGGGLGSQGT--SGRGLGGGGA---GAAAAAAGGAGGG 232
Db 468 GQGAGAAAAAGGAGGGYGGGLGSQAGGGYGGGLGSQAGRGGGGAGAAAAAGGAGGG 527
QY 233 GYGGLGSQGTSGRGLGGGAGAGAAA-----AAAAAGGAGGG 271
Db 528 GYGGLGSQ-AGRGGLGGGAGAAAAAGGAGGGGGLGSQAGGAGAGAAAAAGGAGGG 586
QY 272 YGGLGSQG 279
Db 587 YGGLGSQG 594

RESULT 7
AAM50042
ID AAM50042 standard; Protein; 630 AA.
XX
XX AAM50042;
XX
XX 18-SEP-2002 (first entry)
XX
XX N. clavipes spidroin synthetic homologue SOL protein.
XX
XX Spidroin; spider; silk; fibre; film; membrane; wound; filter; SOL.
XX
XX Synthetic.
XX
XX DE10113781-A1.
XX
XX 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
XX 09-JUN-2000; 2000DE-1028212.
XX
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
XX Scheller J, Conrad U, Grosse F, Guehrs K;
XX
XX WPI; 2002-123561/17.
XX
XX N-PSDB; ABL61043.
XX
XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX wounds, comprises modules that encode repeating units of spidroin
XX proteins
XX
XX Claim 22; Page 48-50; 88pp; German.
XX
XX This invention describes a novel DNA sequence, encoding a synthetic
XX spider silk protein, comprising modules, each comprising a group of
XX sequentially arranged oligonucleotides, each oligonucleotide encoding
XX a repeating unit of a spidroin protein. The synthetic protein has at
XX least 84% homology with the Nephila clavipes spidroin protein and is used
XX to produce synthetic fibres, films and/or membranes, particularly: (i)
XX for medical use, especially to close wounds and/or to support or cover
XX artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX (iii) as filters. The synthetic proteins are very similar to native
XX spider silk proteins; can be prepared on a large scale and can be spun to
XX fibres with excellent mechanical properties (strength and elasticity).
XX Also they retain water solubility after long-term boiling in aqueous
XX solutions and since they are also soluble in organic solvents but
XX precipitated at high salt concentration, they are easily extracted and
XX purified. The modular construction of the invention facilitates
XX incorporation of additional peptide-encoding sequences, e.g. to simplify
XX purification or modulate solubility. This sequence represents the

```

```
CC synthetic N. clavipes spidroin-1 homologue S01 described in the
CC invention.
XX
SQ Sequence 630 AA;

Query Match 69.5%; Score 1031.5; DB 23; Length 630;
Best Local Similarity 59.8%; Pred. No. 1.3e-72;
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;

QY 11 GSMASGRGLGGQGAGAAAA-----AAAAAGGAGGGY 44
  || :|||||
Db 79 GSGQAGRGGLGGQGAGAAAAAGGAGGGYGLGSGQAGRGGQAGAGAAAAAGGAGGGY 138
QY 45 GGLGSGGTSGRGGLGGQGAGAAAAA-----AGGAGGGYGLGSGGTSGRG 95
  ||||| :|||||
Db 139 GGLGSGQ-AGRGGLGGQGAGAAAAAGGAGGGYGLGGQAGGGYGLGSGQ-AGRG 196
QY 96 LGQGQAGAAAA-----AAAAAGGAGGGYGLGSGGTSGRGGLG 135
  |||||
Db 197 LGQGQAGAAAAAGGAGGGYGLGGQAGQGAGAGAAAAAGGAGGGYGLGSGQ-AGRG 252
QY 136 GQGAGAAAAA-----AGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA- 182
  |||||
Db 253 GQGAGAAAAAGGAGGGYGLGGQAGGGYGLGSGQ-AGRGGLGGQGAGAAAAAGG 311
QY 183 -----AAAAAGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA----- 220
  |||||
Db 312 AGGGLGGQGAGQGAGAAAAAGGAGGGYGLGSGQ-AGRGGLGGQGAGAAAAAGGAG 370
QY 221 -----AAAAAGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA- 258
  |||||
Db 371 QGGYGLGSGQAGRGQGAGAAAAAGGAGGGYGLGSGQ-AGRGGLGGQGAGAAAA 429
QY 259 -----AAAAAGGAGGGYGLGSGQ 279
  |||||
Db 430 GGAGGGYGLGSGQAGRGQGAGAAAAAGGAGGGYGLGSGQ 474

RESULT 8
AAM50047
ID AAM50047 standard; Protein; 676 AA.
XX
AC AAM50047;
XX
DT 18-SEP-2002 (first entry)
XX
DE N. clavipes spidroin synthetic homologue S01 protein #2.
XX
KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; S01.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..28
FT Protein /label= LeB4_signal_peptide
FT /note= "synthetic spidroin homologue S01"
FT Region 650..672
FT /note= "c-myc-tag"
FT Domain 673..676
FT /note= "ER retention signal"
XX
DE10113781-A1.
PN
XX
XX 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
XX 09-JUN-2000; 2000DE-1028212.
XX
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
```

```
PI Scheller J, Conrad U, Grosse F, Guehrs K;
XX WPI; 2002-123561/17.
XX
XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX wounds, comprises modules that encode repeating units of spidroin
XX proteins
PS Example 1; Fig 10A; 88pp; German.
XX
XX This invention describes a novel DNA sequence, encoding a synthetic
XX spider silk protein, comprising modules, each comprising a group of
XX sequentially arranged oligonucleotides, each oligonucleotide encoding
XX a repeating unit of a spidroin protein. The synthetic protein has at
XX least 84% homology with the Nephila clavipes spidroin protein and is used
XX to produce synthetic fibres, films and/or membranes, particularly: (1)
XX for medical use, especially to close wounds and/or to support or cover
XX artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX (iii) as filters. The synthetic proteins are very similar to native
XX spider silk proteins; can be prepared on a large scale and can be spun to
XX fibres with excellent mechanical properties (strength and elasticity).
XX Also they retain water solubility after long-term boiling in aqueous
XX solutions and since they are also soluble in organic solvents but
XX precipitated at high salt concentration, they are easily extracted and
XX purified. The modular construction of the invention facilitates
XX incorporation of additional peptide-encoding sequences, e.g. to simplify
XX purification or modulate solubility. This sequence represents a construct
XX composed of the LeB4 signal peptide, N. clavipes spidroin-1 synthetic
XX homologue S01, a c-myc-tag and an endoplasmic reticulum (ER)-retention
XX signal described in the invention.
XX
SQ Sequence 676 AA;

Query Match 69.5%; Score 1031.5; DB 23; Length 676;
Best Local Similarity 59.8%; Pred. No. 1.4e-72;
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;

QY 11 GSMASGRGLGGQGAGAAAA-----AAAAAGGAGGGY 44
  || :|||||
Db 107 GSGQAGRGGLGGQGAGAAAAAGGAGGGYGLGSGQAGRGGQAGAGAAAAAGGAGGGY 166
QY 45 GGLGSGGTSGRGGLGGQGAGAAAAA-----AGGAGGGYGLGSGGTSGRG 95
  ||||| :|||||
Db 167 GGLGSGQ-AGRGGLGGQGAGAAAAAGGAGGGYGLGGQAGGGYGLGSGQ-AGRG 224
QY 96 LGQGQAGAAAA-----AAAAAGGAGGGYGLGSGGTSGRGGLG 135
  |||||
Db 225 LGQGQAGAAAAAGGAGGGYGLGGQAGQGAGAGAAAAAGGAGGGYGLGSGQ-AGRG 280
QY 136 GQGAGAAAAA-----AGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA- 182
  |||||
Db 281 GQGAGAAAAAGGAGGGYGLGGQAGGGYGLGSGQ-AGRGGLGGQGAGAAAAAGG 339
QY 183 -----AAAAAGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA----- 220
  |||||
Db 340 AGGGLGGQGAGQGAGAAAAAGGAGGGYGLGSGQ-AGRGGLGGQGAGAAAAAGGAG 398
QY 221 -----AAAAAGGAGGGYGLGSGGTSGRGGLGGQGAGAAAA- 258
  |||||
Db 399 QGGYGLGSGQAGRGQGAGAAAAAGGAGGGYGLGSGQ-AGRGGLGGQGAGAAAA 457
QY 259 -----AAAAAGGAGGGYGLGSGQ 279
  |||||
Db 458 GGAGGGYGLGSGQAGRGGQAGAGAAAAAGGAGGGYGLGSGQ 502

RESULT 9
AAM50037
ID AAM50037 standard; Protein; 1255 AA.
XX
AC AAM50037;
XX
DT 18-SEP-2002 (first entry)
```


QY 11 GSMASGRGLGGGAGAGAAA-----AAAAAAAAAGGAGGGY 44
 || :|||||
 Db 79 GSQAGRGGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 138
 QY 45 GGLSGQTSRGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 95
 ||||| :|||||
 Db 139 GGLSGQTSRGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 196
 QY 96 LGCGGAGAAA-----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 135
 ||||| :|||||
 Db 197 LGCGGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 252
 QY 136 GCGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 182
 ||||| :|||||
 Db 253 GCGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 311
 QY 183 -----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 220
 ||||| :|||||
 Db 312 AGCGGLGGGAGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 370
 QY 221 -----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 258
 ||||| :|||||
 Db 371 QCGYGGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 429
 QY 259 -----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 279
 ||||| :|||||
 Db 430 GGAGCGGGYGGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 474

RESULT 11
 AAW27178
 ID AAW27178 standard; Protein; 646 AA.
 AC
 AC AAW27178;
 DT
 DT 09-DEC-1997 (first entry)
 XX
 XX Nephila clavipes spider silk protein.
 DE High strength film; fibre; woven article; parachutes; sails;
 KW absorber; body armour; heavy metal; biological weapon; chemical;
 KW flavour; fragrance; Nephila clavipes.
 XX
 OS Nephila clavipes.
 XX WO9708315-A1.
 PN
 PD
 PD 06-MAR-1997.
 PF
 PF 22-AUG-1996; . 96WO-US13767.
 XX
 PR 22-AUG-1995; 95US-0517694.
 XX
 PA (BASE/) BASEL R M.
 PA (ELIO/) ELION G R.
 XX
 PI Basel RM, Elion GR;
 XX WPI; 1997-179272/16.
 DR N-PSDB; AAT85356.
 XX
 PT New opt. multimerised DNA sequences encoding spider silk protein -
 PT cong. both repetitive and non-repetitive sequences, useful for
 PT making high strength films, fibres, woven articles etc.
 XX
 PS Example 2; Fig 1; 57pp; English.
 CC A process has been developed for the production of a DNA fragment
 CC encoding silk protein. The process involves: (a) selecting target
 CC from a silk-producing spider, that contains many repetitive and non-
 CC repetitive regions; (b) selecting a single-stranded DNA primer of at
 CC least 10 nucleotides with a sequence that is complementary to a region
 CC of the target; (c) repetitively combining the primer with melted target

CC DNA, incubating the mixture with nucleotides and a DNA polymerase with
 CC proofreading activity to produce a DNA fragment which is complementary
 CC to the target and is at least 2 kb long. The present sequence encodes
 CC the spider silk protein from Nephila clavipes. The DNA fragment can be
 CC used to make fibres, films, woven articles, e.g. for use in parachutes,
 CC sails, body armour, and absorbers (e.g. of heavy metals, biological
 CC weapons, DNA, chemicals, flavours and fragrances). The high molecular
 CC weight (90-250 kD) of spider silk proteins can be produced on a
 CC commercial scale (at over 2 g/l cell mass). It has better tensile
 CC strength and elasticity than silkworm silk. Inclusion of both repetitive
 CC and non-repetitive regions ensures isolation of stable clones.

XX SQ Sequence 646 AA;

Query Match 69.4%; Score 1031; DB 18; Length 646;
 Best Local Similarity 55.5%; Pred. No. 1.4e-72;
 Matches 244; Conservative 6; Mismatches 10; Indels 180; Gaps 13;

QY 11 GSMASGRGLGGGAGAGAAA-----AAAAAAAAAGGAGGGYCGGLGSG 49
 || :|||||
 Db 50 GSQAGRGGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 109
 QY 50 QGT-----SRRGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 68
 || :|||||
 Db 110 QGAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 169
 QY 69 -----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 109
 ||||| :|||||
 Db 170 AGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 224
 QY 110 AAGAGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 159
 ||||| :|||||
 Db 225 AAGAGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 283
 QY 160 GLGSGQTSRGLGGGAGAGAAA-----SRRGLGGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 182
 ||||| :|||||
 Db 284 GLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 342
 QY 183 -----AAAAAAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 225
 ||||| :|||||
 Db 343 LGNQGAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 401
 QY 226 -----AGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 259
 ||||| :|||||
 Db 402 GYGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 460
 QY 260 AAGAGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 279
 ||||| :|||||
 Db 461 AAGAGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGYCGGLGSGQAGRGGAGAGAAAAGGAGGGY 480

RESULT 12
 AAY40097
 ID AAY40097 standard; protein; 651 AA.

AC AAY40097;
 XX
 XX 19-NOV-1999 (first entry)
 DT
 DT Spider silk protein spidroine major 1.
 XX
 XX Spider silk protein; spidroine major 1; cosmetic; make-up;
 KW dermatological compositions; hair care; skin care; sunscreen;
 KW hormone; moisturizer; skin disorder; skin disorder.
 XX
 OS Nephila clavipes.
 XX
 PN FR2774588-A1.
 XX
 PD 13-AUG-1999.
 XX
 XX 11-FEB-1998; 98FR-0001614.
 PF
 XX

This Page Blank (uspto)

[illegible]

```

Db   468  GCGAGAAAAAAGGAGGGYGGGLSQSGAGGGYGGLGSQCGAGRGCGGAGAAAAAAGAGGAGQG 527
Qy   233  GYGGLSGQTSGRGLGGQGAGAAA-----AAAAAAAAAGGAGQG 271
      ||||| :|||||:|||||:|||||:|||||:|||||:|||||
Db   528  GYGGLSQG-AGRGLGGGAGAAAAAAAGGAGGGGLGSQCGAGAAAAAAGAGGAGQG 586
Qy   272  YGGLGSQG 279
      |||||
Db   587  YGGLGSQG 594

RESULT 3
US-08-556-978B-19
; Sequence 19, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TYPE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
us-08-556-978B-19

Query Match          68.4%; Score 1015; DB 4; Length 651;
Best Local Similarity 59.2%; Pred. No. 3e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps

Qy   11  GSMAASGRGGLGGCGACAGAAA-----AAAAAGGAGGGGYG 45
      || :|||:|||||:|||||:|||||:|||||:|||||
Db   93  GSQAGRGGLGGCGAGAAAAAAAGGAGGGYGGLGNQAGRCQGAAAAAGAGGAGQG 152
Qy   46  GLGSQGTSGRGLGGQGAGAAAAAAA-----AGGAGGGYGGLGSQGTSGRGL 96
      ||||| :|||||:|||||:|||||:|||||:|||||:|||||
Db   153  GLGSQG-AGRGLGGGAGAAAAAAGGAGGGYGGLGGCGAGGGYGGLGSQ-AGRGL 210
Qy   97  GGCGAGAAAAA-----AAAAGGAGGGYGGLGSQGTSGRGLG 135
      |||||:|||||
Db   211  GGCGAGAAAAAAGGAGGGYGGLGGCGAGAGASAAAAGGAGGGYGGLGSQ-----ACRG 266
Qy   136  GQGAGAAAAAAA-----AGGAGGGYGGLGSQGTSGRGLGGCGACAGAAA---- 182

```



```

Query Match          37.4%; Score 555.5; DB 10; Length 529;
Best Local Similarity 44.1%; Pred. No. 8e-28;
Matches 167; Conservative 11; Mismatches 90; Indels 111; Gaps 18;

QY 11 GSMASGRGGLGGCGAG-----AAAAAAAAAG---GAGG---GGYG----- 45
DB 2 GGYPGQGGPGGYGPGQGGPSGSAANANANANGPGGYGPGQGGPGYGGQGGPGRY 61
QY 46 GLGSQGTSGRG-----GLGGQGAG-----AAAAA 74
DB 62 GPGQGGPSGSAANANANGSGQGGPGGYPRQGGPGYGQGGPGSGPGSAANASAA 121
QY 75 GGACQGGYG-----GLGSQGTSGRG-----GLGGQGAGANANANANGAGGQGGYG 121
DB 122 ESGPGGYGPGQGGPGGYGPGQGGPGYGPGQGGPGYGPQGGPGSAANANANGSGQGGPGYG 181
QY 122 GLGSQGTSGRG-----GLGGQGAGANANANANGAGGQGGYGGLGSQGTSGRG---GLG 173
DB 182 P-GQGGPGGYGPGQGGPSGSAANANANGSGQGGPGYG-P-GQGGPGYGPGQGGLS 239
QY 174 GQGAAGAAA-----AAAAAAAAAGGAGGQGGYG-GLGSGQGTSGRG 208
DB 240 GPGSAANANANGPGQGGPGYGPGQGGPSGSAANANANANGPGGYGPGQGGPGYGPG 299
QY 209 GLGGQGAGANANANANGAGGQGGYGGLGSQGTSGRG-----GLGGQGAGANANANANG 264
DB 300 QGGSPGAGSAANANANGPGQGGPGYG-P-QGGPGGYGPGQGGPGYGPGSAANANANG 358
QY 265 -GGAGQGGYGGLGSQGTSG 282
DB 359 PGQGGPGYG-P-GQGGPSG 376

RESULT 6
US-09-861-597-10
; Sequence 10, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861.597
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-861-597-10

Query Match          35.3%; Score 524; DB 10; Length 714;
Best Local Similarity 49.2%; Pred. No. 8e-26;
Matches 155; Conservative 12; Mismatches 107; Indels 41; Gaps 17;

QY 10 HGSMASGRGGLG-GQ----GAGAAAAAANAAG----GAGQGGYGGLGSQGTSGRGLG- 59
DB 296 YGPGQGGPGGYGPGQGGPSGSAANANANGPGGYGPGQGGPGYG-PQGGPGGYGP 354
QY 60 GQGAGANANANANGAGGQGGYGGLGSQGTSGRGLG-GQGAGANANANANGAGAGQG 118
DB 355 GQGGPSGSAANANANGPGQGGYG-PGQGGPGYGPGQGGPSGSAANANANGPGG 413

```


OTHER INFORMATION: Gly at positions 105-201 may be present or absent
US-09-848-990-22

Query Match 29.2%; Score 433; DB 10; Length 201;
Best Local Similarity 42.5%; Pred. No. 1.1e-20;
Matches 110; Conservative 0; Mismatches 91; Indels 58; Gaps

Qy	16	GRGLGGOGAGAAAAAAGACAGGCGYGGILGSQTSRGGILGGOGAGAAAAAAAG	75
Db	1	GGGGGGGGG-----GGGGGGGGGGGGGGGGGGGG-----G	40
Qy	76	GAGCGYGGILGSQTSRGLGGOGAGAAAAAAAAGCAGCGYGGILGSQTSRGGIG	135
Db	41	GGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGG	90
Qy	136	GQGAGAAAAAAGACAGCGYGGILGSQTSRGGILGGOGAGAAAAAAAGCAGCGG	195
Db	91	GGGG-----GGGGPGGGGGGGGGGGGGGGGG-----GGGGGG	131
Qy	196	YGLGSQTSRGLGGOGAGAAAAAAAAGCAGCGYGGILGSQTSRGGILGGOGAGA	255
Db	132	GGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGGGGG	182
Qy	256	AAAAAAGCAGCGYGG	274
Db	183	GGGGGGGGGGGGGGGG	201

RESULT 9

```

US-09-760-364-14
; Sequence 14, Application US/09760364
; Patent No. US20020152479A1
; GENERAL INFORMATION:
; APPLICANT: Lehmann, Juergen Michael
; APPLICANT: Shiau, Andrew Kwan-Nan
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: CAR Modulators: Screening and Treatment of
; TITLE OF INVENTION: Hypercholesterolemia
; FILE REFERENCE: 018781-004110US
; CURRENT APPLICATION NUMBER: US/09/760,364
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/176,398
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (1)..(97)
; OTHER INFORMATION: Gly residues from positions 1-97 may be present or
; OTHER INFORMATION: absent
; NAME/KEY: MOD_RES
; LOCATION: (105)..(201)
; OTHER INFORMATION: Gly residues from, positions 105-201 may be
; OTHER INFORMATION: present or absent
US-09-760-364-14

```

Query Match 29.2%; Score 433; DB 10; Length 201;
Best Local Similarity 42.5%; Pred. NO. 1.1e-20;
Matches 110; Conservative 0; Mismatches 91; Indels 58; Gaps 6;

QY	16	GRGLGGCGAGAAAAAAGGAGCGGGYGGLGSQGTSGRGLGGCGAGAAAAAAGAAG	75
Db	1	GCGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGGGGGGGGGGG--G	40
QY	76	GAGCGGYGGLGSQGTSGRGLGGCGAGAAAAAAGAGCGGYGGLGSQGTSGRGLG	135
Db	41	GCGGGGGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	90

[illegible]

RESULT 10

```

US-09-798-584-18
; Sequence 18, Application US/09798584
; Patent No. US20020102676A1
; GENERAL INFORMATION:
; APPLICANT: Mu, David
; APPLICANT: Powers, Scott
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: KCNB: A NO. US20020102676A1el Potassium Channel Protein
; FILE REFERENCE: 018781-004010US
; CURRENT APPLICATION NUMBER: US/09798,584
; CURRENT FILING DATE: 2001-03-03
; PRIOR APPLICATION NUMBER: US 60/186,951
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly-Gly linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-798-584-18

```

Query Match	29.1%	Score 431.5;	DB 10;	Length 200;
Best Local Similarity	42.5%	Pred. No. 1.3e-20;		
Matches 110; Conservative	0;	Mismatches 90;	Indels 59;	Gaps 6;

QY 16 GRGGLGGQGAGAAAAAAGGAGGGYGGLCSQGTSGRGGGLGGQAGAAAAAAG 75
| | | | | | | | | | | | | | | |
Dd 1 GGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGGGGGGGG-----G 40

[illegible]

QY 136 GQNGAATAAATAAGGAGGCGGCGGCGTSCRGGLGCGGAGATAAATAAGGAGG 195

Dβ 91 GGGG-----GG 130

[illegible]

Qv 256 AAAAAAAAAAGGAGGGYGG 274

pD
182 GGGGGGGGGGGGGGGG 200

RESULT 11

US-09-967-624-19
; Sequence 19, Application US/03967624
; Patent No. US2002014325A1
; GENERAL INFORMATION:
; APPLICANT: Liao, X. Charlene
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge


```
Db          95 YGGLGSQ 101
|||||
RESULT 14
US-09-861-597-5
; Sequence 5, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861.597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-5

Query Match          27.7%   Score 412;   DB 10;   Length 101;
Best Local Similarity 73.6%;   Pred. No. 1.2e-19;
Matches 92;   Conservative 1;   Mismatches 6;   Indels 26;   Gaps 5;

QY 152 GAGGGYGGGSGQTSRGGGGLGGGAGAGAAAAAGAGGAGGGYGGGSGTSGRGGG 211
|||||
Db 3 GAGGGYGGGSGQ-AGRGGGAGAG--AAAAAGGAGCQ-----GGG 44

QY 212 GQAGAGAAAAAGAGGAGGGYGGGSGTSGRGGGAGAGAAAAAGAGGAGGG 271
|||||
Db 45 SQAGAGGAGAAAAAGGAGGGYGGGSGQ-----AGRGGAG-----AAAAAGGAGGG 96

QY 272 YGGLG 276
|||||
Db 97 YGGLG 101

RESULT 15
US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
```

;

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063120

;

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063121

;

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486

;

PRIOR FILING DATE: 1997-10-21

PRIOR APPLICATION NUMBER: 60/063540

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063541

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063544

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063564

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063734

;

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063870

;

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/064103

;

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/065311

;

PRIOR FILING DATE: 1997-11-13

PRIOR APPLICATION NUMBER: 60/066120

;

PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: 60/066466

;

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066772

;

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/069335

;

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069425

;

PRIOR FILING DATE: 1997-12-12

PRIOR APPLICATION NUMBER: 60/069870

;

PRIOR FILING DATE: 1997-12-17

PRIOR APPLICATION NUMBER: 60/068017

;

PRIOR FILING DATE: 1997-12-18

PRIOR APPLICATION NUMBER: 60/077450

;

PRIOR FILING DATE: 1998-03-10

PRIOR APPLICATION NUMBER: 60/077632

;

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/077649

;

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/078886

;

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/078939

;

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/079664

;

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079786

;

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/080107

;

PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080194

;

PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080327

;

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080333

;

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/081049

;

PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081070

;

PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081195

;

PRIOR FILING DATE: 1998-04-09

PRIOR APPLICATION NUMBER: 60/081838

;

PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/082568

;

PRIOR FILING DATE: 1998-04-21

PRIOR APPLICATION NUMBER: 60/082569

;

PRIOR FILING DATE: 1998-04-21

PRIOR APPLICATION NUMBER: 60/082704

;

PRIOR FILING DATE: 1998-04-22

;	PRIOR APPLICATION NUMBER: 60/082790	;
;	PRIOR FILING DATE: 1998-04-22	;
;	PRIOR APPLICATION NUMBER: 60/083322	;
;	PRIOR FILING DATE: 1998-04-28	;
;	PRIOR APPLICATION NUMBER: 60/083495	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083496	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083499	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083559	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/084366	;
;	PRIOR FILING DATE: 1998-05-05	;
;	PRIOR APPLICATION NUMBER: 60/084414	;
;	PRIOR FILING DATE: 1998-05-06	;
;	PRIOR APPLICATION NUMBER: 60/084639	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/084640	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/084643	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/085573	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085579	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085580	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085582	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085700	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/086023	;
;	PRIOR FILING DATE: 1998-05-18	;
;	PRIOR APPLICATION NUMBER: 60/086392	;
;	PRIOR FILING DATE: 1998-05-22	;
;	PRIOR APPLICATION NUMBER: 60/086486	;
;	PRIOR FILING DATE: 1998-05-22	;
;	PRIOR APPLICATION NUMBER: 60/087098	;
;	PRIOR FILING DATE: 1998-05-28	;
;	PRIOR APPLICATION NUMBER: 60/087208	;
;	PRIOR FILING DATE: 1998-05-28	;
;	PRIOR APPLICATION NUMBER: 60/087609	;
;	PRIOR FILING DATE: 1998-06-02	;
;	PRIOR APPLICATION NUMBER: 60/087759	;
;	PRIOR FILING DATE: 1998-06-02	;
;	PRIOR APPLICATION NUMBER: 60/087827	;
;	PRIOR FILING DATE: 1998-06-03	;
;	PRIOR APPLICATION NUMBER: 60/088025	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088028	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088029	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088033	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088167	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088202	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088212	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088217	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088326	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088655	;
;	PRIOR FILING DATE: 1998-06-09	;
;	PRIOR APPLICATION NUMBER: 60/088722	;
;	PRIOR FILING DATE: 1998-06-10	;
;	PRIOR APPLICATION NUMBER: 60/088738	;
;	PRIOR FILING DATE: 1998-06-10	;
;	PRIOR APPLICATION NUMBER: 60/088740	;

2	PRIOR FILING DATE:	1998-06-10
3	PRIOR APPLICATION NUMBER:	60/008811
4	PRIOR FILING DATE:	1998-06-10
5	PRIOR APPLICATION NUMBER:	60/008824
6	PRIOR FILING DATE:	1998-06-10
7	PRIOR APPLICATION NUMBER:	60/008825
8	PRIOR FILING DATE:	1998-06-10
9	PRIOR APPLICATION NUMBER:	60/008826
10	PRIOR FILING DATE:	1998-06-10
11	PRIOR APPLICATION NUMBER:	60/008861
12	PRIOR FILING DATE:	1998-06-11
13	PRIOR APPLICATION NUMBER:	60/008863
14	PRIOR FILING DATE:	1998-06-11
15	PRIOR APPLICATION NUMBER:	60/008876
16	PRIOR FILING DATE:	1998-06-11
17	PRIOR APPLICATION NUMBER:	60/008909
18	PRIOR FILING DATE:	1998-06-12
19	PRIOR APPLICATION NUMBER:	60/0089105
20	PRIOR FILING DATE:	1998-06-12
21	PRIOR APPLICATION NUMBER:	60/008912
22	PRIOR FILING DATE:	1998-06-16
23	PRIOR APPLICATION NUMBER:	60/0089514
24	PRIOR FILING DATE:	1998-06-16
25	PRIOR APPLICATION NUMBER:	60/0089538
26	PRIOR FILING DATE:	1998-06-17
27	PRIOR APPLICATION NUMBER:	60/0089598
28	PRIOR FILING DATE:	1998-06-17
29	PRIOR APPLICATION NUMBER:	60/0089653
30	PRIOR FILING DATE:	1998-06-17
31	PRIOR APPLICATION NUMBER:	60/0089508

Query Match	27.4%	Score 407.5;	DB 12;	Length 1894;
Best Local Similarity	35.0%;	Pred. No. 1.8e-18;		
Matches 133; Conservative	12;	Mismatches 122;	Indels 113;	Gaps 13;

Search completed: December 18, 2002, 16:38:50
Job time : 8.01995 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 8.59824 Seconds
(without alignments)
3253.588 Million cell updates/sec

Title: US-09-490-291-2
Perfect score: 1485
Sequence: 1 MRGSHHHHHGSMASGRGGL.....YGLGSGQGTSGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1015	68.4	718	2 A36068	major ampullate fi
2	617.5	41.6	2639	2 T31328	fibroin - Chinese
3	610.5	41.1	641	1 Q0BE31	nuclear antigen EB
4	556.5	37.5	1901	2 F70806	hypothetical glyci
5	556	37.4	1489	2 D70807	hypothetical glyci
6	554	37.3	627	2 A4112	spidroin 2, dragli
7	544.5	36.7	1381	2 E70806	hypothetical glyci
8	543.5	36.6	1079	2 E70807	hypothetical glyci
9	538	36.2	1538	2 H70846	hypothetical glyci
10	535	36.0	749	2 A70812	hypothetical glyci
11	525.5	35.4	1660	2 A70869	hypothetical glyci
12	525	35.4	384	1 A26099	glycine-rich cell
13	519.5	35.0	853	2 A70896	hypothetical glyci
14	516.5	34.8	778	2 F70963	hypothetical glyci
15	514	34.6	1306	2 A70934	hypothetical glyci
16	512.5	34.5	767	2 E70895	hypothetical glyci
17	512	34.5	484	2 E70846	hypothetical glyci
18	507.5	34.2	923	2 E70820	hypothetical glyci
19	503	33.9	491	2 D70916	hypothetical glyci
20	503	33.9	731	2 C70974	hypothetical glyci
21	499	33.6	714	2 A70807	hypothetical glyci
22	498.5	33.6	882	2 B70812	hypothetical glyci
23	496.5	33.4	465	1 S01820	glycine-rich cell
24	496	33.4	603	2 A70770	hypothetical glyci
25	495	33.3	576	2 A70900	hypothetical glyci
26	495	33.3	1329	2 E70917	hypothetical glyci
27	494	33.3	667	2 A70893	hypothetical glyci
28	494	33.3	914	2 H70987	hypothetical glyci
29	490	33.0	338	1 KNMU	glycine-rich cell

30	490	33.0	801	2 F70824	hypothetical glyci
31	489.5	33.0	291	1 S31415	glycine-rich prote
32	489.5	33.0	396	2 T49109	glycine-rich prote
33	489.5	33.0	618	2 A70989	hypothetical glyci
34	489	32.9	498	2 C70720	hypothetical glyci
35	487.5	32.8	694	2 F70868	hypothetical glyci
36	487	32.8	496	2 H70839	hypothetical glyci
37	486.5	32.8	741	2 G70917	hypothetical glyci
38	486	32.7	1011	2 F70620	hypothetical glyci
39	484	32.6	615	2 H70589	hypothetical glyci
40	481	32.4	591	2 B70523	hypothetical glyci
41	480.5	32.4	532	2 F70580	hypothetical glyci
42	480.5	32.4	783	2 E70824	hypothetical glyci
43	479	32.3	594	2 G70545	hypothetical glyci
44	477.5	32.2	606	2 H70816	hypothetical glyci
45	477	32.1	837	2 E70835	hypothetical glyci

ALIGNMENTS

RESULT 1
A36068

major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)
C:Species: Nephila clavipes
C:Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997
C:Accession: A36068
R:Xu, M.; Lewis, R.V.
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990
A:Title: Structure of a protein superfiber: spider dragline silk.
A:Reference number: A36068; MUID:90384959; PMID:2402494
A:Accession: A36068
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-718 <XUA>
A:Cross-references: GB:M37137; NID:g159711; PID:g159712
A:Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue

Query Match 68.4%; Score 1015; DB 2; Length 718;
Best Local Similarity 59.2%; Pred. No. 1.9e-53;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

Qy	11	GSMASGRGLGGGAGAGAAA-----	-----AAAAAGGAGGCGYG	45
Db	93	GSQAGRGRLGGGAGAGAAAAAAGAGGCGYGGGLGNQAGRGCGGCAAAAAGGAGCGGCGY	152	
Qy	46	GLGSGQGTSGRGLGGGCGAGAGAAAAA-----	-----AGGAGCGGCGYGGGLGSGQTSGRGGL	96
Db	153	GLGSGQ-AGRGGLGGGAGAGAAAAAGGAGGCGYGGGLGGGAGGCGYGGGLGSGQ-AGRGGL	210	
Qy	97	GCQAGAGAAAA-----	-----AAAGGAGCGGCGYGGGLGSGQTSGRGGLG	135
Db	211	GCQAGAGAAAAAGGAGCGGGLGGGCGAGAGAGAGGAGGCGYGGGLGSGQ-AGRGGL	266	
Qy	136	GQAGAGAAAAA-----	-----AGGAGCGGCGYGGGLGSGQTSGRGGLGGGAGNAAA-----	182
Db	267	GEGAGAAAAAGGAGCGGCGYGGGLGGGAGGCGGCGYGGGLGSGQ-AGRGGLGGGAGAGAAAGGAG	325	
Qy	183	-----	-----AAAAAGGAGGCGYGGGLGSGQTSGRGGLGGGAGAGAAAA-----	220
Db	326	QGGLGGGCGAGAGAGAGAGAGGCGGCGYGGGLGSGQ-AGRGGLGGGAGAGAGAGAGAGAGAG	384	
Qy	221	-----	-----AAAAAGGAGGCGYGGGLGSGQTSGRGGLGGGAGAGAAAA-----	259
Db	385	GGYGGGLGSGAGRGCGGAG	443	
Qy	260	-----	-----AAAAAGGAGGCGYGGGLGSGQ	279
Db	444	GGAGCGGCGYGGGLGNQAGRGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	485	

RESULT 2
T31328

A;Contents: annotation; protein coding region

Db	536	ACGAGGAGGNTG	VGTTNC	SGGQG	-GTGAGGAGGAGG	VGADNPTG	IGTGGTGGKG	-GA	592
Qy	170	GGLGGQG	--AGAAAAA	AAAGG	-AGQGGYGL	SGSQGT	---SCRG	GLGGQGA	AAAAAA
Db	593	GGAGGQGS	GAGTNC	SGCAGCTG	GQGGAGGAGG	AGADNPTG	IGCAGCTG	TGGGAAG	652
Qy	224	AAAGGAGCG	YGG	-----	LG	SQGT	SGRGLGGQ	-CAGAAAAA	AAAAA
Db	653	GAGGAGTGTG	TGAV	SVGNAG	IGTGGTGG	VGAGGAGG	AAAAA	AAAGS	TGGGAGFAGG
Qy	272	YGLG	SGQTS	GI	283				
Db	713	EGGAG	--GNSGV	722					

RESULT 5
D70807
hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, K.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, R.; A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G. A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome A: Reference number: A70500; PMID:98295987; PMID:9634230

A: Accession: D70607
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA

A; molecule type: DNA
A; Residues: 1-1489 <COL>
A; Cross-references: GB:AL020222; GB:AL123456; NID:g3361554; PIDN:CAA17751.1; PID:g292445
A; Experimental source: strain H37Rv
C; Genetics:
A; Gene: RV3514
C; Superfamily: collagen alpha 1(IV) chain

Query Match	37.4%	Score 556;	DB 2;	Length 1489;
Best Local Similarity	39.3%;	Pred. No. 3.5e-26;		
Matches 140;	Conservative 20;	Mismatches 109;	Indels 87;	Gaps 12;

Qy 15 SRGGGLGQGAGAAAAAAGAGAGGGYG-LSGGTSGRGLGG-----GAGAA 66
: ||| | : ||| | : ||| | : ||| | : ||| |
Db 457 TGTGGTGGDGGTGAACAGAGAGGAGTAGTGCTGMIGTTNAGVVGAGGGGGDGGAGGAGAD 516

QY 67 A-----AAAAAGGA-----GQGGYGGLGSQG-----TSCR 93

Db 517 ADQPGATGGTGAGGAGGAGGSSGAGGTNGSGGAGTGGCGGAGGAGGAGADNPtGI 576

F_A 57 GGTGGGCGCTGGAACAGGAGCAAGTGGTGGMIGTTGNAGVGGCAGGCGGGGAGSAGADADO 636
Db 577 GGTGGGCGCTGGAACAGGAGCAAGTGGTGGMIGTTGNAGVGGCAGGCGGGGAGSAGADADO 636

QY 147 -----AAAAGGAGQ-----GGYGGLGSQGTSGRGGL 172

Db 637 PGATCGTGFAAGAGGAGCAGAGGSSSAGGTSSGSAGGTCRQSGTGGAGGAGADNP TGIGGT 696

[illegible]

Qy 232 GYGGLGSQT---SGRGLGGQAGAAAAAAGGAGQGGYGG-LGSQGTSGI 283

Db 757 GGAGGAGGAGADNPTGIGTGGDGTGGAAGAGGAGGAGTGGTGMIGTTGNAGV 812

RESULT 6
MA112

spidron 2, dragline silk fibroin - orb spider (*Nephila clavipes*) (fragment)
N: Alternate names: silk fibroin, dragline
C: Species: *Nephila clavipes*

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998
C:Accession: A44112; S27824
R:Hanman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A:Title: Isolation of a clone encoding a second dragline silk fibroin. Neph
A:Reference number: A44112; MUID:92406076; PMID:1527052
A:Accession: A44112
A:Status: preliminary
A:Molecule type: mrna
A:Residues: 1-627 <H1N>
A:Cross-references: GB:M92913; NID:q159713; PID:q159714
A:Note: Sequence extracted from NCBI backbone (NCBI:P:113893)
R:Hanman, M.B.; Lewis, R.V.
Submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin
A:Reference number: S27824
A:Accession: S27824
A:Molecule type: mrna
A:Residues: 19-627 <H12>
A:Cross-references: EMBL:M92913

Query Match 37.3%; Score 554; DB 2; Length 627;
Best Local Similarity 44.1%; Pred. No. 2.5e-26;
Matches 165; Conservative 15; Mismatches 101; Indels 94; Gaps

QY

3 GSHHHHHCMSACRGGLGGCAGAAAAAAAGGACG---GGYG-----GLGSQ 50
| : | | | : |||||::: | |||| | |
D6

87 GGCGPRCGPGGYGGCGOOGPSGPSSAAAASNAASASSGOOPGGYPGEOOGPCGYGSGOO 146

QY 51 GTSGRG----GLGCGCAGAAAAAAGAGCGGCGGLSGCTSGRG---CLGGCGAG 102

Db 147 GPGGYGPGGCGSPGSAASAAAAAAGCPGCGGGYGP-GGCGPGGYCGGCGCPGCGSA 205

QY 103 AAAAAAAAAAGGACGGCGGYGGLGSQTSGRG-----GLGGQGAGAAAA- 144
||||||| | G |||| | || | || | : ||||
db 206 AAAAAAASGPQOOCPPGGYGP-GQOCPPGGYGPQOOISLGPSSAAAAAAGAPGOOCPPGGYGP 25A

QY	145	-----AAAAAGGACGGYG-GLGSGQTSGRGLGGCAGAAAAAAGGAGQ	193
Dh	285	GOOGSPGSGSAAAAAAGCGGCGPGOOGSGGCGSAAAAAAGCGGOGL	324

QY	194	GGVGGGSGQTSGRG---	GLGGGAGAAAAA-	GGAGGGVGGGSGQTSGR--	246
			:		
Dh	325	GGVGP-GGGPGGVPGCGGCPGCSA	AAACACCCGCGVCP-CGCCGCGCSA	382	

Qy 247 -----GLGGGAG-----AAAAAAAAAAG---GAGQGSGYGG 274
| | | | : |||||
ph 383 SAAAAAAACBCGCVCBCCGCCGCVABCCGCCGCCSAAAAAAAACVCCCAACCACCCG 412

QY 275 L--GSQGTSGIRPAA 288
| | | |
Dh 413 VAPCCCGTCGCGCAAN 459

RESULT 7
F70805

hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change:20-Jun-2000

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N. Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton S.; Squares,

A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.
A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the com
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>

A; Cross references: GB:AL022022; GB:AL123430; FIDN:CAA17744.1
A; Experimental source: strain H37Rv

C:Genetics:
A:Gene: RV3507
C:Superfamily: collagen alpha 1(IV) chain

Query Match 36.7%; Score 544.5; DB 2; Length 1381;
Best Local Similarity 37.8%; Pred. No. 1.6e-25;
Matches 144; Conservative 17; Mismatches 115; Indels 105; Gaps 13;

QY 1 MRGSHHHHIGSMASRGRLGGGAGAGAAAAAAGAGAGGGYGL--GSQGTSGRGL 58
DB 205 MGGTGGNGGNGALLIGGGGLG--GAGMGCTGGTGTGTGCGNGGALLIGAGVGGAGGI 262
QY 59 GGGAGAGAAAAAAGAGAGGGYGLGGGAGAGAAAAAAGAGAGAGAAAAAAGCA--- 115
DB 263 GGCTGAGGAAGAGGTG--GNGGAGGLFMNG--GDGAGGGGGDGAAGDAASAGGTGK 318
QY 116 -----GQGGYGLGSGQ--GTSGRGGLGGQ-----AG 140
DB 319 GGGGGGCTGGAGGAPVLFHCGAGMGCGGTGCGAGGAGGDTTVIAAGTGEGGTGG 378
QY 141 AAAAAAAGAGAGQGG--YGLGSGQTSGRGLGGGAGAGAAAAAAG-----GAGQG 194
DB 379 AAGAGGAAGARGALTSGGLAGGVGAGTGTGTCTGGNGADAAAVVFGANGDPGFAGKG 438
QY 195 GYGLGSGQ-----GTSGRGGLGQ-----CAGAAAAA 221
DB 439 GNGGIGAAVTVGVAGDGGTGGKGTGGAGGAGNDAGSTGNPGKGDDGIGGAGGAGGA 498
QY 222 AAAAA--AGGAGGGYGLGSGQTSGRGLGGQ-----GAGAAAAA 261
DB 499 AGTGNGGHAGTGDGGGTGGNGGTGGVNCADNTLNPTDTPGAGGPGGAGGAGG 558
QY 262 AAGGAGGGYGLGSGQTS 282
DB 559 AAGGPGGTGGTGGNGGNG 579

RESULT 8

B70807
Hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>
A:Cross-references: GB:AL020222; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3512
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 36.6%; Score 543.5; DB 2; Length 1079;
Best Local Similarity 40.7%; Pred. No. 1.5e-25;
Matches 144; Conservative 12; Mismatches 109; Indels 89; Gaps 14;

QY 14 ASGRGLGGQGA-GAAAAAAGAGAGGGYGLGSGQTS-----GRG 56
DB 217 ANGIAGNGDGDGAAGAVGIGSATCAGDGHGHTGAAGNGGTGAGGSGIDGVGGGTG 276
QY 57 GLGGGAGAAAAAAGGAG--QGGYGLGSGQ-----GTSGRGGLGGQ-- 100
DB 277 GTGGNGGGAIGAGGADAGSGNSGNGGIGKGKNAGAGGAGNSGTVGANGTGGDGG 336
QY 101 -AGAAAAAAGAGAGGGYGLGSGQTSGRGLGGQ-----AGAAAAA 148

DB 337 NGGAAGAATAGSNGGAGTGSAGGNG--GTGRRGSGGAGGDTGGVGGKGNCADEVG 394
QY 149 AAGGAG-----QGGYGLGSGQTSGRGLGGGAGAGAAAAAAGAGAGGGYGL-- 199
DB 395 GAGAGGSGPNTSPFGNGGGGCGGGGGAG-GAAGAGGAGGANGTAGNCGGGAGGTGG 453
QY 200 -----GSOQTSGRGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSGRG 246
DB 454 AGAASATNNGSGGAGGTGGDGGSGGAGTGGAGGTGGAAGDGGGQGGGAGG--GAGGQG 512
QY 247 GLGCGQ-----AGAAAAAAGCA--GCGY--GGLGSGQTS 282
DB 513 GAGGAGGTGCGNITGCTAGTAGAAGNGCAACKGGAGGCGGTGGTGGGGGAG 566

RESULT 9

H70846
Hypothetical glycine-rich protein RV345c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1538 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g326
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV345c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 36.2%; Score 538; DB 2; Length 1538;
Best Local Similarity 38.9%; Pred. No. 4.2e-25;
Matches 126; Conservative 13; Mismatches 131; Indels 54; Gaps 9;

QY 11 GSMASRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSGRGLGGGAGAGAAAAA 70
DB 1030 GMINGLGGFGGAGGAVDVAATTGAGNGGAGGFASTGLGPGGAGGPGGAGFASG 1089
QY 71 AAGAGCA---GQGYGGLGSGQ-----TSGRGLGGGAGAGAAAAAAG----- 114
DB 1090 VGGVGGAGDGGAGGAGGAGGAGGAGTGGGEGRTGGNGSGDGGGIGLGGNGGLGNGGVS 1149
QY 115 -----AGQGGYGLGSGQTSGRGLGGGAGAGAAAAAAGAGAGGGYGLGSGQ--- 164
DB 1150 ETGFGGAGNGYGGPG--GPGENGGLGGNGGAGGNGGVSTTTGGDGGAGGKGGNGDGGN 1207
QY 165 -----GTSRGLGGGAGAGAAAAAAGG-----AGQGGYGLGSGQTS 205
DB 1208 VGLGGDAGSGGAGNGGIGTDTAGGAGGAGGAGGNGSSKSTTTGNAGSGGAGNGGTGLN 1267
QY 206 GRGGLGGGAGAGAAAAA---AAGCAGCGGYGLGSGQTS--GRGGLGGGAGAGAAAAA 261
DB 1268 GAGGAGGAGGAGVAGVSGFNAVGGDGNCGNGCHGDDCTTGGAGGKGGNGSAGSG 1327
QY 262 ---AAGGAGGGYGLGSGQTS 282
DB 1328 VYVNTAGHCGNGNGNGNGSAG 1351

RESULT 10

A70812
Hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: A70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-749 <COL>
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261507; PIDN:CAAL17639.1; PID:g291689
A;Experimental source: strain H37Rv
C;Genetics:
C;Gene: Rv0833
C;Superfamily: elastin

Query Match 36.0%; Score 535; DB 2; Length 749;
Best Local Similarity 39.9%; Pred. No. 3.7e-25;
Matches 133; Conservative 9; Mismatches 123; Indels 68; Gaps 11;

Qy 16 GRGGLGGGAGAAAAAAGAGGAGGCGGGLGSGQ-----GTSGR 55
Db 3 GNGGAGSGAPCAICGAGGPAGLIGVGAGGAGGDSAVAGVIGGAGGAGGAALLFGAGGA 62

Qy 56 GGLGGGAGAAAAAAGGA-----GCGGYGGLGSGQTSRGRGLGGGAGAAAAA 108
Db 63 GGAGGSG--GSGAGGAGGAGGAGGLFASGSGCGFCGFASTGTGGAGGTGGAGGLFASGGV 121

Qy 109 AAAAGGAGGCGGGLGSGQ-----TSRGRGLGGGAGGAAAAAAGGAGGCGG 161
Db 122 GGTGGAGSGTGTGGTGGAGGLFASGAGGAGGSGTGTGAGTGGAGGLFAGGAGGL 181

Qy 162 GSQG--TSRGRGLGGQ-----GAGAAAAAAGGA--GCGGYGGL--GSQGTSGR 207
Db 182 GGQGNHTGCGGAGGSAGLLALGDDGAGGAGGAATTGTGGAGGAGGAGLLFGSGGAGGS 241

Qy 208 GGLGGQ-----GAGAAAAAAGGAGGCGGGLGSGQTSRGRGLGGGAGGA---- 255
Db 242 GGAAGTGTGTCNSGAGGAGGAGKALLFGSGGAGGSGGAGGFANGTGGAGGAGGLIG 301

Qy 256 -----AAAAAAGGAGGCGGGLGSGQTSGI 283
Db 302 NGNGSGSGTVAITGGAGNGGAGGAG--GGAGL 332

RESULT 11
A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70869
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70869
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1660 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAAL16067.1; PID:g279153
A;Experimental source: strain H37Rv
C;Genetics:
C;Gene: Rv2490c
C;Superfamily: collagen alpha 1(IV) chain

Query Match 35.4%; Score 525.5; DB 2; Length 1660;
Best Local Similarity 40.7%; Pred. No. 2.4e-24;
Matches 137; Conservative 12; Mismatches 117; Indels 71; Gaps 13;

Qy 14 ASRGRGLGGQ--GAGAAAAAAGGAGGAGCGGYGGLGSGQTSGR-----GGLGGGGA 63
Db 1022 AGGAGGAGGAGGAGGVSVDGGAGCGNCGAGCGGAGGAGGANGIDISGTGGAGG 1081

Qy 64 GAAAAAAGGAGGAGCGGYG--LGSQGTSGRGLGGGAGAAAAAAGG--- 114
Db 1082 GGGDGGAGGCGHGGDGGVGAAPSGTGVSHGTGGVGGDGLG--GAGGCGAGGNGGIGI 1140

Qy 115 --AGCGYGGGLGSG--GTSRGRGLGGQ-----GAGAAAAA 147
Db 1141 TVGAGGAGGNGDPPGAGGRGLGDSNGTSAANGVDASKHGLPLTGGDGVGGNGAKAA 1200

Qy 148 AAGGAGCGGYGGLGSGQTSRGRGLGGGAGAAAAAAGGA--GCGGYG--LGSQ 202
Db 1201 AAGDGGCGCGD--GNAGLFGDGGAGGCGADGTAALGDDGGAGGAGKGDAGDIDG 1258

Qy 203 GTSRGRGLGGGA-----GAAAAAAGGA--GCGGYGGLGSGQTSRGL-- 248
Db 1259 GDGKGGDGAHGLGLTVAGNGGAGGAGGAGGAGGAGGAGFLDGGNGGAGCGGAGRGSP 1318

Qy 249 ---GGGAGAAAAAAGGAGGCGGYGGLGSGQTSG 282
Db 1319 CGCGVCGHGGCAGGDAGMNGGGTGGCGGNGAAGGAG 1355

RESULT 12
A26099
glycine-rich cell wall structural protein - garden petunia
C;Species: Petunia x hybrida (garden petunia)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A26099
R;Condit, C.M.; Meagher, R.B.
Nature 323, 178-181, 1986
A;Title: A gene encoding a novel glycine-rich structural protein of petunia.
A;Reference number: A26099
A;Accession: A26099
A;Molecule type: mRNA
A;Residues: 1-384 <CON>
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 35.4%; Score 525; DB 1; Length 384;
Best Local Similarity 43.9%; Pred. No. 9e-25;
Matches 125; Conservative 2; Mismatches 136; Indels 22; Gaps 5;

Qy 12 SMASRGRGLGGGAGAAAAAAGAGCGGYGGLGSGQTSRGRGLGGGAGAAAAA 71
Db 60 SFRGRGAGCGFGGAGGAGGGLGGGGGAG--GGGLGGGGAGGFGGAGG 117

Qy 72 AAAGG--AGCGYGGGLGSGQTSRGRGLG-----GQAGAAAAAAGGAGG 119
Db 118 CAGGGLGGGGLGGGGGAGGCGGCGGAGGSGGFGAGGCGGAGAGGCGGFGG 177

Qy 120 YGGLGSGQTSRGRGLGGGAGAAAAAAGGAGCGGYGGLG--GTSRGG--- 171
Db 178 GGGGCGGSGHGGFGGAGGCGGGLGGGCGGCGGCGGCGGCGGCGGCGGCGG 237

Qy 172 LGGCGAGAAAAAAGGAGCGGYGGLGSGQTSRGRGLGGGAGAAAAAAGGAGCAG 231
Db 238 CGGVGGVGGGAAGGGGGGGGGGGLGGSGHGGGFGAGGCGGAGGCGGAGG 297

Qy 232 GGYGGLGSGQTSRGRGLGGGAGAAAAAAGGAGCGGYGGLG 276
Db 298 GGGGCGGCGGSGHGGFGAGGCGGCGGAGGGLGGGGGAGGGGGG 342

RESULT 13
A70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C;Accession: A70896
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Thu Dec 19 09:28:10 2002

Db 1201 GFGBGLDASSG 1212

Search completed: December 18, 2002, 16:21:57
Job time : 12.5982 secs

This Page Blank (uspto)

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 4.46662 Seconds
(without alignments)
2702.183 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485

Sequence: 1 MRGSHHHHHGSMASGRGGL.....YGLGSGQTGIRPAKLN 291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	68.4	747	1 SPDL_NEPCL	P19837 nephila cla
2	610.5	41.1	641	1 EBN1_EBV	P03211 epstein-bar
3	556.5	37.5	1901	1 Y208_MYCTU	O53553 mycobacteri
4	554	37.3	627	1 SPDL_NEPCL	P46804 nephila cla
5	525	35.4	384	1 GRP1_PETHY	P09789 petunia hyb
6	516.5	34.8	778	1 Y034_MYCTU	P71933 mycobacteri
7	508.5	34.2	5263	1 FBOH_BOMMO	P05790 bombyx mori
8	496.5	33.4	465	1 GRP2_PHAVU	P10496 phaseolus v
9	496	33.4	603	1 YD25_MYCTU	O10637 mycobacteri
10	494	33.3	914	1 WA22_MYCTU	O06794 mycobacteri
11	490	33.0	338	1 GRP_ARATH	P27483 arabidopsis
12	490	33.0	801	1 Y747_MYCTU	O53810 mycobacteri
13	489	32.9	498	1 Y118_MYCTU	O50615 mycobacteri
14	472	31.8	860	1 ELS_MOUSE	P54320 mus musculu
15	467	31.4	957	1 Y278_MYCTU	P56877 mycobacteri
16	463	31.2	543	1 YP91_MYCTU	Q50630 mycobacteri
17	453.5	30.5	864	1 ELS_RAT	P18165 mus musculu
18	444.5	29.9	463	1 Y468_MYCTU	P04985 bos taurus
19	443.5	29.9	491	1 YK98_MYCTU	Q53416 mycobacteri
20	438	29.5	252	1 GRP1_PHAVU	Q10707 mycobacteri
21	418.5	28.2	515	1 Y140_MYCTU	P10495 phaseolus v
22	409.5	27.6	481	1 LORI_MOUSE	O50594 mycobacteri
23	394.5	26.6	747	1 ELS_BOVIN	P18165 mus musculu
24	385.5	26.0	750	1 ELS_CHICK	P04985 bos taurus
25	378	25.5	316	1 LORI_HUMAN	P23490 homo sapien
26	374	25.2	730	1 ELS_HUMAN	P15502 homo sapien
27	363	24.4	375	1 SANT_PLAFV	P09593 plasmodium
28	357.5	24.1	2038	1 FSH_DROME	P13709 drosophila
29	341.5	23.0	401	1 CSP_PLACC	P08674 plasmodium
30	332.5	22.4	1733	1 VNVA_PRYKA	P33485 pseudorabie
31	330.5	22.3	1380	1 DDX9_MOUSE	O70133 mus musculu
32	328	22.1	398	1 CSP_PLACC	P08673 plasmodium
33	325	21.9	1356	1 CA21_ONCMY	O93484 oncorhynch

ALIGNMENTS

RESULT 1

ID	SPDL_NEPCL	STANDARD	PRT	747 AA
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwitt R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-1- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M37137; AAA29380.1; -;			
DR	EMBL; U03848; AAB60212.1; -;			
DR	PIR; A36068; A36068.			
DR	Silk; Repeat.			
FT	NON_TER	1	655	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	25	1.
FT	REPEAT	26	38	2.
FT	REPEAT	39	66	3.
FT	REPEAT	67	96	4.
FT	REPEAT	97	130	5.
FT	REPEAT	131	158	6.


```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z80225; CAB02341.1; -
DR EMBL; AE007103; AAK47026.1; ALT_INIT.
DR TIGR; MT2712; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 34.88; Score 516.5; DB 1; Length 778;
Best Local Similarity 42.18; Pred. No. 1.6e-19;
Matches 136; Conservative 9; Mismatches 122; Indels 56; Gaps 16;

QY 14 ASGRGGLGGGGA--GAAAAAAGAGG-----AGGGVGGGLSGGTS-----GR 55
Db 454 AGGAGGAGGIGTGGAGGATGEGGAGTGVGGVGGAGGGGGGHTGVDGDFGGD 513

QY 56 GGLGGGAGAAAAA--AAAGGAGAGGGYGG--LGSGTSGRGLGGGGAGAAA--AAAA 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 GGICGDCNGALGAAGNGTGGAGNGGRCGMLGNGGAGGAGTGTGGGGAAGFAGGV 573

QY 110 AAAGGAG-----QGGYGLGS-QTSGRGLGGGGA-----GAAAAAAGAGGAG 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 GGAGGEGTLDGAGTAEGGTGGLGGLGGVGTGGMGGSGVGGNGGAAGSLGLGGGG 633

QY 157 GYGLGSGQTSRGLGGQ---CAGAAAAAAGAGAGGGYGGTSGTSGRGLGG 212
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 GVGGTG--GTGGGAGGNGGAGGAGTTGGGTATGGGGTGGVGGAG--GTGTGGAGG 689

QY 213 Q--GAGAAA-----AAAAAGGAGGGYGLGGTSGRGLGGGGAGAAAAAAG 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 TTGGGCGAGGLGAGAGGAGTGGTGGGGLGGQ--GNGNGGTGATGGGQGGDFALGG 747

QY 266 GAGGGYGL--GSQTSGRTRP 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 NGGAGGAGGSPGSSGIGQNMGP 770

RESULT 7
FBOH_BOMMO STANDARD; PRT; 5263 AA.
AC P05790; Q26379; Q17220;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FIBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
```

```
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions.";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions.";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kinshu X Showa;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RX STRAIN=J-139;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Blochim. Biophys. Acta 1432:92-103(1999).
CC -1- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -1- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -1- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF226688; AAF76983.1; -
DR EMBL; V00094; CAA23432.1; -
DR EMBL; V00097; CAA23433.1; -
DR EMBL; S74439; AAB31861.1; -
DR EMBL; X13869; CAA32076.1; -
DR EMBL; M35378; AAA27839.1; -
DR EMBL; AB017362; BAA33147.1; -
DR PIR; S01844; S01844.
KW Silk; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 5263
FT DOMAIN 149 5206
FT DISULFID 5244 5244
FT DISULFID 5260 5263
FT CONFLICT 10 10
C -> V (IN REF. 2).
```


RESULT 12
Y747_MYCTU
ID Y747_MYCTU STANDARD; PRF; 801 AA.
AC OS3810;
CD 30-MAY-2000 (Rel. 39, Created)
DT DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE Hypothetical PE-PGRS family protein Rv0747 precursor
GN GN RV0747 OR MT0772.5 OR MV041.21.
OS OS Mycobacterium tuberculosis.
OC OC Bacteria; Actinobacteria; Actinobacteria (class); Actino-
CC CC Actinomycetales; Corynebacterineae; Mycobacteriaceae
OX OX NCBI_TaxId=1773;
[1]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=H37Rv;
RC RC MEDLINE=98295987; PubMed=9634230;
RA RA Cole S.T., Brosch R., Parkhill J., Garnier T., Church-
RA RA Gordon K.V., Eiglmeier K., Gas S., Barry C.E. III, Ten
RA RA Badcock S.K., Basham D., Brown D., Chillingworth T., Co-
RA RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin
RA RA Hornsbury T., Jagels K., Krogh A., McLean J., Moule S.
RA RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Ra-
RA RA Rutter S., Seeger K., Skelton S., Squares S., Squares
RA RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT RT "Deciphering the biology of Mycobacterium tuberculosis
RL RL complete genome sequence.";
Nature 393:537-544(1998).
[2]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=CDC 1551 / Oshkosh;
RA RA Fieischmann R.D., Allard D., Eisen J.A., Carpenter L.
RA RA Peterson J.J., DeBoy R., Dodson R., Gwinn M.L., Haft D.
RA RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D.
RA RA Delcher A., Utterback T., Weidman J., Khouri H., Gill
RA RA Bisbal W.;
"Whole genome comparison of Mycobacterium tuberculosis
laboratory strains.";
Submitted (APR-2001) to the EMBL/GenBank/DBJ database
-!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY
SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced thro-
between the Swiss Institute of Bioinformatics and t-
the European Bioinformatics Institute. There are no
use by non-profit institutions as long as its con-
modified and this statement is not removed. Usage b-
entities requires a license agreement (See http://www.e-
or send an email to license@sib-swiss.ch).

EMBL; AL021958; CAAL7514.1; -
EMBL; AE006968; AAK45011.1; -
TIGR; MT0772.5; -
TubercuList; Rv0747; -
InterPro: IPR000084; PE_region.
Pfam; PF009334; P5; 1.
Hypothetical protein; Repeat; Signal; Complete proteo-
POTENTIAL.
SIGNAL 1 30
CHAIN 31 801
FT FT
CONFLICT 188 188 G -> S (IN REF. 2).
CONFLICT 225 225 R -> G (IN REF. 2).
CONFLICT 227 227 R -> G (IN REF. 2).
CONFLICT 295 295 K -> R (IN REF. 2).
CONFLICT 300 300 S -> G (IN REF. 2).
CONFLICT 338 338 T -> I (IN REF. 2).
CONFLICT 377 377 A -> P (IN REF. 2).
CONFLICT 577 577 T -> A (IN REF. 2).
CONFLICT 580 595 MISSING (IN REF. 2).
SEQUENCE 801 AA; 65407 MW; EA54C9BF45A00F41 CRC64

33.08; Score 490; DB 1; Length 801;

Best Local Similarity 38.6%; Pred. No. 3.3e-18;
Matches 133; Conservative 14; Mismatches 120; Indels 78; Gaps 14;

QY 11 GSMASGRGLGCGGCAAAAAAAGG-----GQGGYGGGLGSGQTSGRGLGGQG--- 62
Db 425 GGLWLFNGNGGQGTIGVNGCAGGAGGAGGILFTGTTGGSGGPGATGLGGIGAGGAA 484
QY 63 -----AGAAAAAAGG-----AGQGGYGGGLGSGQTSGRGLGGQGCGAGA 103
Db 485 LLFGSGGAGGGGAGAVGGNGAGGAGNAGALLGAAGAGAGGAGAVG--GNGGAGNG--GL 541
QY 104 AAAAAAAGC----AQGGYGGGLGSG-QTSGRGLGGCGAGAAAAAAGG-----AG 154
Db 542 FANGCAGGPGGFGPAGAGGTGGAGNGGLFGAGGTGGAGGTGGAGGAGGAGGGLFG 601
QY 155 QGGYGGGLGSGQTS-----GRGLGGQ-----GAGAAAAAAGG----- 191
Db 602 AGTGTGAGSHSTACVSGGAGGAGGAGDAGLLSLGASGAGGSGGSLTAAVVGIGAGG 661
QY 192 ---GQGGYGGGLGSGQTSGRGLGGQGAGA-----AAAAAAGGAGGQGGYGGGLGSGQ 241
Db 662 LLFGSGGAGGSGGFSNSGNGGAGGAGGAGDAGLLVSGGAGGAGASATGAATGGDGAG--G 719
QY 242 TSGRGGYGGCGAGAAAAAAGGAGGAGGQGGYGG-----LGSQGTSG 282
Db 720 KSGAFGLGGDGGAGGATGLSGAFHTGGKGGVGGSAVLINGNGG 764

RESULT 13
Y118_MYCTU STANDARD; PRT; 498 AA.
ID Y118_MYCTU AC Q50615;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV1818c.
GN RV1818C OR MT1866 OR MTCYI11.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagals J., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann J.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

[illegible]

Qy	174	GCGAGAAAAAAGGAGGGYGGGLGSQTSG-----RGLGGGAGAGAAAAA	222
Db	746	GTGAAGSGSGITTTGGGHGNAGLLSLGASGAGGSGGASSLGGAGTGTGGNALLFG	805
Qy	223	AAAAGGAG-----CCGYGGGLGSQ-----GTSGRGGLGGGAGAAAAAAAGG	266
Db	806	FRGAGGAGHGHAALTTCQGAGGAGNGGGLLFSGAGAGGAGGSCAN---ALGACTGT	862
Qy	267	AGCGGYGGL-GSQGTSGIRR	285
Db	863	GGDGGHAGVFCNGDGDCRR	892

Search completed: December 18, 2002, 16:17:54
Job time : 9.46662 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 17.0848 seconds
(without alignments)
3509.535 Million cell updates/sec

Title: US-09-490-291-2
Perfect score: 1485
Sequence: 1 MRGSHHHHHGSMASRGGL.....YGLGSGQTSIGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP-archaea.*
- 2: SP-bacteria.*
- 3: SP-fungi.*
- 4: SP-human.*
- 5: SP-invertebrate.*
- 6: SP-mammal.*
- 7: SP-mhc.*
- 8: SP-organelle.*
- 9: SP-phage.*
- 10: SP-plant.*
- 11: SP-rodent.*
- 12: SP-virus.*
- 13: SP-vertebrate.*
- 14: SP-unclassified.*
- 15: SP-virus.*
- 16: SP-bacteriap.*
- 17: SP-archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	68.7	617	5	O46172
2	1017	68.5	644	5	Q8WSW4
3	1008	67.9	544	5	O46171
4	818.5	55.1	447	5	O9BIT1
5	817.5	55.1	648	5	O9BIT7
6	768.5	51.8	691	5	O9BIU3
7	689.5	46.4	854	5	O9BIU4
8	680	45.8	360	5	O9BIU0
9	675	45.5	988	5	O17434
10	640.5	43.1	410	5	O16988
11	634	42.7	253	5	O9BIT4
12	632.5	42.6	373	5	O9BIT9
13	617.5	41.6	2639	5	O76786
14	602	40.5	2655	5	O964F4
15	597.5	40.2	233	5	O9BIT6
16	593.5	40.0	738	5	O02402

17	567.5	38.2	399	5	O9BIT8
18	565.5	38.1	563	5	O9BIT5
19	563	37.9	1217	16	O8VIY9
20	563	37.9	1953	5	O9BIT7
21	560.5	37.7	436	5	O967T8
22	557	37.5	1715	16	O8VI20
23	556	37.4	1489	16	O53559
24	545	36.7	444	5	O9BIU6
25	544.5	36.7	1381	16	O53552
26	543.5	36.6	1079	16	O53557
27	541.5	36.5	294	5	O16986
28	540	36.4	1384	16	O8VI21
29	538	36.2	1507	16	O8VI23
30	538	36.2	1538	16	O53395
31	535	36.0	749	16	O53844
32	526.5	35.5	421	5	O93119
33	525.5	35.4	1665	16	O53215
34	524.5	35.3	390	10	O9M3V2
35	520.5	35.1	619	12	O9IP09
36	519.5	35.0	853	16	O53439
37	514	34.6	636	5	O16987
38	514	34.6	1306	16	O53775
39	512.5	34.5	767	16	O53435
40	512	34.5	484	16	O53394
41	509.5	34.3	1468	5	O9GUB5
42	507.5	34.2	923	16	O53890
43	507	34.1	610	5	O9V5V8
44	506	34.1	342	5	O9BIU1
45	503	33.9	491	16	O06818

ALIGNMENTS

RESULT 1

O46172 PRELIMINARY; PRT; 617 AA.
ID O46172
AC O46172
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Dragline silk protein spidroin 1 (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
RT proteins (spidroin) from Nephila clavipes (Tetragnathidae) and Araneus
RT bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98318805;
RA Beckwith R., Arcidiacono S., Stote R.;
RT "Evolution of repetitive proteins: spider silks from Nephila clavipes
RT (Tetragnathidae) and Araneus bicentenarius (Araneidae).";
RL Insect Biochem. Mol. Biol. 28:121-130(1998).
DR EMBL; U37520; AAC04504.1; -;
FT NON_TER 1
FT NON_TER 617
SQ SEQUENCE 617 AA; 49410 MW; 1EFOCE9269A832E2 CRC64;

Query Match 68.7%; Score 1020; DB 5; Length 617;

Best Local Similarity 55.3%; Pred. No. 4.8e-53;

Matches 242; Conservative 6; Mismatches 12; Indels 178; Gaps 13;

OY 11 GSMASRGGLGGGAGAGAAA-----AAAAAAAAAGGAGGCGGGLGGGCGGAGGAGGAGGCGGGLGS 49

DB 47 GSOGAGRGGLGGGAGAGAAA-----AAAAAAAAAGGAGGCGGGLGS 106

[illegible]

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 45.4697 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-4
Perfect score: 3493
Sequence: 1 MRGSHHHHHHGMASGRGGL.....GGYGPQQQTSGIRRPAAKLN 681

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3493	100.0	681	22	AAB82609
2	3390	97.1	691	22	AAB82610
3	2419	69.3	1255	23	AAW50037
4	2419	69.3	1880	23	AAW50039
5	2372	67.9	989	23	AAW50038
6	2294	65.7	1617	23	AAU11794
7	2294	65.7	1626	23	AAU11798
8	2254	64.5	818	23	AAU11797
9	2250	64.4	809	23	AAU11793
10	2139	61.2	630	23	AAW50042

11	2139	61.2	676	23	AAW50047	N. clavipes spidro
12	2135	61.1	651	20	AAV40097	Spider silk proteol
13	2135	61.1	651	23	AAU11781	Spider natural sil
14	2135	61.1	718	19	AAW53346	Nephila clavipes s
15	2135	61.1	718	21	AAV59070	N. clavipes spider
16	2126	60.9	718	12	AAW14308	N. clavipes draglin
17	2076	59.4	606	16	AAW99053	Spider dragline va
18	2076	59.4	606	20	AAV40100	Polymer of an anal
19	2052.5	58.8	646	18	AAW27178	Nephila clavipes s
20	1972.5	56.5	606	16	AAW99055	Spider dragline va
21	1972.5	56.5	606	20	AAV40102	Polymer of an anal
22	1971.5	56.4	604	16	AAW99057	Spider dragline va
23	1963.5	56.2	606	20	AAV40101	Polymer of an anal
24	1711.5	49.0	528	22	AAW82611	Spider recombinant
25	1609.5	46.1	773	23	AAW50045	N. clavipes spidro
26	1609.5	46.1	777	23	AAW50046	N. clavipes spidro
27	1458.5	41.8	364	23	AAW50043	N. clavipes spidro
28	1425	40.8	291	12	AAW82608	Spider recombinant
29	1405.5	40.2	831	16	AAW80168	pMIS1 Misp spider
30	1395	39.9	615	20	AAV40099	Spider silk proteol
31	1387	39.7	655	23	AAW76672	Protein related to
32	1331.5	38.1	714	16	AAW99059	Spider dragline va
33	1331.5	38.1	714	20	AAV40103	Polymer of an anal
34	1233.5	35.3	1177	9	AAW80940	SLPIII protein com
35	1233.5	35.3	1177	11	AAW05307	SLP III (Silk-fibr
36	1233.5	35.3	1177	17	AAW95105	Silk like protein
37	1233.5	35.3	1177	18	AAW26342	Silk-like protein
38	1233.5	35.3	1177	21	AAW78277	SLPIII amino acid
39	1233.5	35.3	1177	22	AAW72725	Repetitive protein
40	1233.5	35.3	1177	22	AAW63995	SLPIII protein seq
41	1233.5	35.3	1178	14	AAW41007	Silk-like protein
42	1233.5	35.3	1178	19	AAW53518	Amino acid sequenc
43	1176.5	33.7	1023	19	AAW53524	Amino acid sequenc
44	1176.5	33.7	1059	14	AAW41013	SLP4 multimeric pr
45	1176.5	33.7	1059	18	AAW26348	SLP4 synthetic pro

ALIGNMENTS

RESULT 1
AAB82609
ID AAB82609 standard; Protein; 681 AA.
XX AAB82609;
XX
XX 02-OCT-2001 (first entry)
DT
DE Spider recombinant silk protein POE((SP1)4/(SP2)1)4.
XX
XX Spider; orb-weaver; silk protein; POE((SP1)4/(SP2)1)4;
KW structural protein; purification; fibre; spinning.
XX
XX Nephila clavipes.
OS
PN WO200153333-A1.
XX
PD 26-JUL-2001.
PF
PF 01-NOV-2000; 2000WO-US30086.
XX
XX 20-JAN-2000; 2000US-0490291.
XX
XX (MELL/) MELLO C M.
PA (ARCI/) ARCIDIACONO S.
PA (BUTL/) BUTLER M M.
PA (USSA) US SEC OF ARMY.
XX
PI Mello CM, Arcidiacono S, Butler MM;
XX WPI; 2001-483136/52.
DR N-PSDB; AAW26302.
XX

QY 632 GAAAAAAGAGAGGCGGGLGSGQTSRGGYGP-GQOTSG 672
 Db 791 G-----AAAAAGAGCGGCGGLGSGQ-AGCGYGGGSGAG 827

RESULT 7

AAU11798

ID AAU11798 standard; Protein; 1626 AA.

XX AAU11798;

XX 26-MAR-2002 (first entry)

XX Dragline protein 1 analogue DP-1B/his tag 16mer.

XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;
 KW fabric production; material construction; rope; surgical suture;
 KW flexible tie down; electrical component; implantation;
 KW Dragline protein 1; DP-1B 16mer; His tag.

XX Nephila clavipes.
 OS Synthetic.
 XX WO200190389-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US16937.

XX 25-MAY-2000; 2000US-206968P.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX Yang JG;

XX WPI; 2002-106209/14.

XX Producing silk-like proteins in a green plant, useful in fabrics and
 PT material construction, comprises providing a silk-like protein
 PT expression cassette to a green plant -
 XX Example 1; Page 87-92; 93pp; English.

XX The invention relates to a method of producing silk-like proteins (SLP)
 CC in a green plant by providing a green plant containing an SLP expression
 CC cassette. The green plant contains the expression cassette P-SLP-T where
 CC P is a promoter for driving the expression of an SLP transgene and T is a
 CC 5' terminator. The method is useful for producing silks and silk-like
 CC proteins in green plants. The silks and silk-like proteins may be used in
 CC fabrics or in material construction, such as rope, surgical sutures,
 CC flexible tie downs for certain electrical components, or as a
 CC biomaterial for implantation. The method allows for more cost effective
 CC production of silk not obtained from natural or microbial sources.
 CC The present sequence is the silk protein dragline protein (DP) 1
 CC synthetic variant DP1B/his tag 16mer (i.e. 16 copies of DP1B with a His
 CC tag at the C-terminus) which is used as an SLP construct in the method of
 CC the invention.

XX Sequence 1626 AA;

Query Match 65.7%; Score 2294; DB 23; Length 1626;
 Best Local Similarity 61.6%; Pred. No. 7.1e-160;
 Matches 543; Conservative 13; Mismatches 38; Indels 288; Gaps 36;

QY 11 GSMASGRGLGGGCGAGAAAA-----AAAAAGAGCGGCGGGLGS 49

Db 16 GSQAGRGGLGGGCGAGAAAAAGGAGCGGGLGSGQAGAGAAAAAGAGCGGCGGGLGS 75

QY 50 QGTSGRGLGGGCGAGAAAAAAGAGCGGCGGGLGSQGT-----SGRGLG 97

Db 76 QG-----AGRGCGAG-----AAAAAGAGCGGCGGGLGSGQAGCGGCGGGLG 127

QY 98 GCGAGAAAA-----AAAAAGAGCGGCGGGLGSGQTSRGGGLG 136

PD 29-NOV-2001.

Db 128 GCGAGAAAAAAGAGCGGGLGSGQAGAGAAAAAGAGCGGCGGGLGSGQ-----ACRGG 183
 QY 137 QGAGAAAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAAAA- 194
 Db 184 QGAG-----AAAAAGAGCGGCGGGLGSGQ-AGCGYGGGSGQAGRGLGGGAGAAAA 238
 QY 195 -----AAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAAAA 234
 Db 239 AAGAGCGGGLGSGQAGAGAAAAAAGAGCGGCGGGLGSGQ-----ACRGGCGAG-----NA 290
 QY 235 AAAAGAGCGGCGGGLGSGQGT-----SGRGLGGGAGAAAA----- 270
 Db 291 AAAAGAGCGGCGGGLGSGQAGCGGCGGGLGSGQAGRGLGGGAGAAAAAAGAGCGGGLG 350
 QY 271 -----AAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAAAAAAGAGCGG 321
 Db 351 SQAGCGAGAAAAAAGAGCGGCGGGLGSGQ-----AGRGCGAG-----AAAAAGAGCGG 402
 QY 322 YGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAAAA----- 359
 Db 403 YGGLGSGQ-AGCGYGGGSGQAGRGLGGGAGAAAAAAGAGCGGGLGSGQAGAGA 461
 QY 360 AAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAAAAAAGAGCGGCGGGLGSGQGT- 418
 Db 462 AAAAGAGCGGCGGGLGSGQ-----AGRGCGAG-----AAAAAGAGCGGCGGGLGSGQAG 513
 QY 419 -----SGRGLGGGCGAGAAAA-----AAAAAGAGCGG 446
 Db 514 QGCGGGLGSGQAGRGLGGGCGAGAAAAAAGAGCGGGLGSGQAGAGAAAAAGAGCGG 573
 QY 447 GYGLGSGQTSRGLGGGCGAGAAAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQ 505
 Db 574 GYGLGSGQ-----AGRGCGAG-----AAAAAGAGCGGCGGGLGSGQ-AGCGYGGGLG 624
 QY 506 TSGRGLGGGCGAGAAAA-----AAAAAGAGCGGCGGGLGSGQGT 544
 Db 625 GAGRGLGGGCGAGAAAAAAGAGCGGCGGGLGSGQAGAGAAAAAAGAGCGGCGGGLGSGQ- 683
 QY 545 SGRGLGGGCGAGAAAAAAGAGCGGCGGGLGSGQGT-----SGRGLGGG 592
 Db 684 ---AGRGCGAG-----AAAAAGAGCGGCGGGLGSGQAGCGGCGGGLGSGG 736
 QY 593 AGAAAA-----AAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSG 631
 Db 737 AGAAAAAAGAGCGGGLGSGQAGAGAGAAAAAAGAGCGGCGGGLGSGQ-----ACRGGCGA 792
 QY 632 GAAAAAAGAGCGGCGGGLGSGQTSRGGYGP-GQOTSG 672
 Db 793 G-----AAAAAGAGCGGCGGGLGSGQ-AGCGYGGGLGSGQAG 829

RESULT 8

AAU11797

ID AAU11797 standard; Protein; 818 AA.

XX AAU11797;

XX 26-MAR-2002 (first entry)

XX Dragline protein 1 analogue DP-1B/his tag 8mer.

XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;
 KW fabric production; material construction; rope; surgical suture;
 KW flexible tie down; electrical component; implantation;
 KW Dragline protein 1; DP-1B 8mer; His tag.

XX Nephila clavipes.
 OS Synthetic.

XX WO200190389-A2.

XX 29-NOV-2001.

XX	24-MAY-2001; 2001WO-US16937.	QY	419	-----SGRGGLGGGAGAAAA-----	446
XX	25-MAY-2000; 2000US-206968P.	Db	514	QGGYGGGLSGOGAGRGGLGGGAGAAAAAAGGAGGGLSGOGAGGAGAAAAAAGGAGG	573
XX	(DUPO) DU PONT DE NEMOURS & CO E I.	QY	447	GYGGLGSGQTSGRGGLGGGAGAAAAAAGGAGGAGGAGGGLSGOGTSGPGGYP-GQ	505
XX	Yang JG;	Db	574	GYGGLGSGQ-----AGRGGQAG-----	624
XX	WPI; 2002-106209/14.	QY	506	TSRGGLGGGAGAAAA-----	544
XX	Producing silk-like proteins in a green plant, useful in fabrics and	Db	625	GAGRGGLGGGAGAAAAAAGGAGGAGGGLSGOGAGGAGAAAAAAGGAGGGLSGQ-	683
XX	material construction, comprises providing a silk-like protein	QY	545	SGRGGLGGGAGAAAAAAGGAGGAGGGLSGOGT-----	592
XX	expression cassette to a green plant -	Db	684	---AGRGGQAG-----	736
XX	Example 1; Page 83-86; 93pp; English.	QY	593	AGAAAA-----	631
XX	The invention relates to a method of producing silk-like proteins (SLP)	Db	737	AGAAAAAAGGAGGGLSGOGAGGAGAAAAAAGGAGGAGGGLSGQ-----	792
XX	in a green plant by providing a green plant containing an SLP expression	QY	632	GAAGAAAAAAGGAGGAGGGLSG	655
XX	cassette. The green plant contains the expression cassette P-SLP-T where	Db	793	G-----	812
XX	P is a promoter for driving the expression of an SLP transgene and T is a	XX	RESULT 9		
XX	5' terminator. The method is useful for producing silks and silk-like	ID	AAU11793		
XX	proteins in green plants. The silks and silk-like proteins may be used in	XX	AAU11793 standard; protein; 809 AA.		
XX	fabrics or in material construction, such as rope, surgical sutures,	AC	AAU11793;		
XX	flexible tie downs for certain electrical components, or as a	DT	26-MAR-2002 (first entry)		
XX	biomaterial for implantation. The method allows for more cost effective	XX	Dragline protein 1 analogue DP-1B 8mer.		
XX	production of silk not obtained from natural or microbial sources.	DE	Orb-weaving spider; silk-like protein; SLP; transgenic plant;		
XX	The present sequence is the silk protein Dragline protein (DP) 1	XX	promoter; 5' terminator; fabric production; material construction; rope;		
XX	synthetic variant DP1B/his tag 8mer (i.e. 8 copies of DP1B with a His tag	KW	surgical suture; flexible tie down; electrical component; implantation;		
XX	at the C-terminus) which is used as an SLP construct in the method of the	KW	Dragline protein 1; DP-1B 8mer.		
XX	invention.	XX	Nephila clavipes.		
XX		OS	Synthetic.		
XX		XX	WO200190389-A2.		
XX		XX	29-NOV-2001.		
XX		XX	24-MAY-2001; 2001WO-US16937.		
XX		XX	25-MAY-2000; 2000US-206968P.		
XX		PA	(DUPO) DU PONT DE NEMOURS & CO E I.		
XX		PI	Yang JG;		
XX		XX	WPI; 2002-106209/14.		
XX		XX	Producing silk-like proteins in a green plant, useful in fabrics and		
XX		XX	material construction, comprises providing a silk-like protein		
XX		XX	expression cassette to a green plant -		
XX		XX	Disclosure; Page 74-76; 93pp; English.		
XX		XX	The invention relates to a method of producing silk-like proteins (SLP)		
XX		XX	in a green plant by providing a green plant containing an SLP expression		
XX		XX	cassette. The green plant contains the expression cassette P-SLP-T where		
XX		XX	P is a promoter for driving the expression of an SLP transgene and T is		
XX		XX	5' terminator. The method is useful for producing silks and silk-like		
XX		XX	proteins in green plants. The silks and silk-like proteins may be used		
XX		XX	in fabrics or in material construction, such as rope, surgical sutures,		
XX		XX	flexible tie downs for certain electrical components, or as a		
XX		XX	biomaterial for implantation. The method allows for more cost effective		
XX		XX	production of silk not obtained from natural or microbial sources.		
XX		XX	The present sequence is the silk protein Dragline protein (DP) 1		
XX		XX	synthetic variant DP1B/his tag 8mer (i.e. 8 copies of DP1B with a His tag		
XX		XX	at the C-terminus) which is used as an SLP construct in the method of the		
XX		XX	invention.		
SQ	Sequence 818 AA;				
	Query Match				
	Best Local Similarity 64.5%; Score 2254; DB 23; Length 818;				
	Matches 533; Conservative 11; Mismatches 34; Indels 286; Gaps 34;				
QY	11 GSMASGRGLGGGAGAAAA-----	QY	11	GSMASGRGLGGGAGAAAA-----	49
Db	16 GSOGAGRGGLGGGAGAAAAAAGGAGGGLSGOGAGGAGGAGGAGGGLSG	Db	16	GSOGAGRGGLGGGAGAAAAAAGGAGGGLSGOGAGGAGGAGGAGGGLSG	75
QY	50 QGTSRGGLGGGAGAAAAAAGGAGGAGGGLSGOGT-----	QY	50	QGTSRGGLGGGAGAAAAAAGGAGGAGGGLSGOGT-----	97
Db	76 QG-----AGRGGQAG-----	Db	76	QG-----AGRGGQAG-----	127
QY	98 GOGAGAAAA-----	QY	98	GOGAGAAAA-----	136
Db	128 GOGAGAAAAAAGGAGGGLSGOGAGGAGAAAAAAGGAGGAGGGLSGQ-----	Db	128	GOGAGAAAAAAGGAGGGLSGOGAGGAGAAAAAAGGAGGAGGGLSGQ-----	183
QY	137 QGAGAAAAAAGGAGGAGGGLSGOGTSGPGGYP-GQOTSGRGGLGGGAGAAAA-	QY	137	QGAGAAAAAAGGAGGAGGGLSGOGTSGPGGYP-GQOTSGRGGLGGGAGAAAA-	194
Db	184 QGAG-----	Db	184	QGAG-----	238
QY	195 -----	QY	195	-----	234
Db	239 AAGGAGGGLSGOGAGGAGAAAAAAGGAGGAGGGLSGQ-----	Db	239	AAGGAGGGLSGOGAGGAGAAAAAAGGAGGAGGGLSGQ-----	290
QY	235 AAAAGGAGGGLSGOGT-----	QY	235	AAAAGGAGGGLSGOGT-----	270
Db	291 AAAAGGAGGGLSGOGGAGGGLSGOGAGGAGGAGAAAAAAGGAGGAGGGLG	Db	291	AAAAGGAGGGLSGOGGAGGGLSGOGAGGAGGAGAAAAAAGGAGGAGGGLG	350
QY	271 -----	QY	271	-----	321
Db	351 SQAGGAGAAAAAAGGAGGAGGGLSGQ-----	Db	351	SQAGGAGAAAAAAGGAGGAGGGLSGQ-----	402
QY	322 YGGLASQTSRGPGGYP-GQOTSGRGGLGGGAGAAAA-----	QY	322	YGGLASQTSRGPGGYP-GQOTSGRGGLGGGAGAAAA-----	359
Db	403 YGGLSGQ-AGOGGYGGLSGOGAGRGGLGGGAGAAAAAAGGAGGGLSGOGAG	Db	403	YGGLSGQ-AGOGGYGGLSGOGAGRGGLGGGAGAAAAAAGGAGGGLSGOGAG	461
QY	360 AAAAGGAGGGLSGOGTSGRGGLGGGAGAAAAAAGGAGGAGGGLSGOGT-	QY	360	AAAAGGAGGGLSGOGTSGRGGLGGGAGAAAAAAGGAGGAGGGLSGOGT-	418
Db	462 AAAAGGAGGGLSGOG-----	Db	462	AAAAGGAGGGLSGOG-----	513

XX Producing silk-like proteins in a green plant, useful in fabrics and
PT material construction, comprises providing a silk-like protein
PT expression cassette to a green plant -
XX
XX
PS Disclosure; Page 69-71; 93pp; English.
XX
CC The invention relates to a method of producing silk-like proteins (SLP)
CC in a green plant by providing a green plant containing an SLP expression
CC cassette. The green plant contains the expression cassette P-SLP-T where
CC P is a promoter for driving the expression of an SLP transgene and T is a
CC 5' terminator. The method is useful for producing silks and silk-like
CC proteins in green plants. The silks and silk-like proteins may be used in
CC fabrics or in material construction, such as rope, surgical sutures,
CC flexible tie downs for certain electrical components, or as a
CC biomaterial for implantation. The method allows for more cost effective
CC production of silk not obtained from natural or microbial sources.
CC The present sequence is orb-weaving spider spidroin1, a natural silk
CC protein whose sequence is used as the basis for the SLP proteins used in
CC the method of the invention.
XX
SQ Sequence 651 AA;

Query Match 61.1%; Score 2135; DB 23; Length 651;
Best Local Similarity 68.3%; Pred. No. 1.4e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;
QY 23 QGAGAAAAAAGAGGAGGGGGLGSGGT--SGRGLGGGAG-AAAAAAAAAGGAGG 79
DB 1 QGAG----AAAAAGAGGGGGLGGGAGGGGGLGGGAGGAGAGAAAAAGAGG 56
QY 80 GYGGLGSGGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGGTSGRGLGGG 139
DB 57 GYGGLGSGG----AGRGGGAG----AAAAAGGAGGGGGLGSGG-AGRGLGG 107
QY 140 GAAAAAAGAGGAGGGGGLGSGGTSGPGGYPGQGTSGRGLGGGAGAAAAA 199
DB 108 G-----AGRGGGAG-----AGRGGG-----AGRGGG-----AGRGG 142
QY 200 AGGAGGGGGLGSGGTSGRGLGGGAGAAAAA-----AGCAGGGGGLG 250
DB 143 AGGAGGGGGLGSGG-ACRGLGGGAGAAAAAAGAGGAGGGGGLGGGAGGG 201
QY 251 SGTSGRGLGGGAGAAAAA-----AAAAAGAGGGGGLG 289
DB 202 SGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGGTSGPGGYP-GQGTSGRGL 260
QY 290 QGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGGTSGPGGYP-GQGTSGRGL 348
DB 261 QG----AGRGGGAG----AAAAAGGAGGGGGLGGG-AGGGGGLGGGAGRGL 311
QY 349 GGGGAGAAAA-----AAAAAGGAGGGGGLGSGGTSGRGLGGG 390
DB 312 GGGGAGAAAAAGGAGGGGGLGGGAGGAGAAAAAAGAGGAGGGGGLGGG 370
QY 391 AGAAAAAAGAGGAGGGGGLGSGGTSGRGLGGGAGAAAAAAGAGGAGGGG 450
DB 371 AG---AVAAAAAGAGGGGGLGSGG-----ACRGGGAG-----AAAAAGGAGGG 419
QY 451 LGSQGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGGT-----SGP 497
DB 420 LGNG-AGRGLGGGAG---AAAAAAGGAGGGGGLGNGAGRGGGAGGAGGAG 475
QY 498 GYGYP-GQGTSGRGLGGGAGAAAAA-----AGGAGGGGGLGSGGTSGRGL 550
DB 476 GYGGLGSGGAGR---GGGAGAAAAAAGAGGAGGGGGLGSGG-AGRGL 531
QY 551 GGGGAGAAAA-----AAAAAGGAGGGGGLGSGGTSGRGLGG 590
DB 532 GGGGAGAAAAAGGAGGGGGLGGGAGGAGAAAAAAGGVRGGGGLGSGG----AGRGG 587
QY 591 QGAGAAAAAAGAGGAGGGGGLGSGGTSGRGLGGGAGAAAAAAGAGGAGGG 650

DB 588 QGAG----AAAAAGGAGGAGGGGGLGGG-GRGGLGGGAG-----AAAAAGGAGGG 636
QY 651 GGLGS 655
DB 637 GGVGS 641
RESULT 14
AAW53346
ID AAW53346 standard; Protein; 718 AA.
XX
AC AAW53346;
XX
DT 06-JUL-1998 (first entry)
XX
DE Nephila clavipes spider silk protein.
XX
KW Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;
KW cocoon; tensile strength; elasticity.
XX
OS Nephila clavipes.
XX
PN US5728810-A.
XX
PD 17-MAR-1998.
XX
PF 19-APR-1995; 95US-0425069.
XX
PR 15-APR-1991; 91US-0684819.
PR 20-APR-1990; 90US-0511792.
PR 04-OCT-1994; 94US-0317844.
PR 19-APR-1995; 95US-0425069.
XX
PA (UYWY-) UNIV WYOMING.
XX
PI Hinman MB, Lewis RV, Xu M;
XX
DR WPI; 1998-270437/24.
DR N-PSDB; AAV23249.
XX
PT Recombinant spider silk proteins - useful for making fibres
XX
PS Claim 1; Column 29-34; 68pp; English.
XX
CC The present sequence represents a spider silk protein from the present
CC invention. Spider silk proteins, and peptide fragments within the
CC proteins, can be produced and purified independently and can then be
CC mixed and made into fibres that have higher tensile strengths and
CC elasticity than naturally occurring fibres. The fibres can be used in
CC mixed composites. The invention allows the two naturally occurring
CC Nephila clavipes silk proteins to be produced independently so that
CC they can later be combined to form silk fibres of high tensile strength
CC and elasticity.
XX
SQ Sequence 718 AA;
Query Match 61.1%; Score 2135; DB 19; Length 718;
Best Local Similarity 68.3%; Pred. No. 1.5e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;
QY 23 QGAGAAAAAAGAGGAGGGGGLGSGGT--SGRGLGGGAG-AAAAAAAAAGGAGG 79
DB 1 QGAG----AAAAAGAGGGGGLGGGAGGGGGLGGGAGGAGAAAAAAGGAGG 56
QY 80 GYGGLGSGGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGGTSGRGLGGG 139
DB 57 GYGGLGSGG----AGRGGGAG----AAAAAGGAGGGGGLGSGG-AGRGLGG 107
QY 140 GAAAAAAGAGGAGGGGGLGSGGTSGPGGYPGQGTSGRGLGGGAGAAAAA 199
DB 108 G-----AGRGGGAG-----AGRGGG-----AGRGGG-----AGRGG 142
QY 200 AGGAGGGGGLGSGGTSGRGLGGGAGAAAAA-----AGGAGGGGGLG 250

Db 143 AGGAGGGYGGGLGSGQ-AGRGGLGQAGAGAAAAAGGAGGGYGGGLGQGGAGGGYGGGLG 101
QY 251 SGTSGRGLGQAGAGAAAAA-----AAAAGAGGGYGGGLG 289
Db 202 SQG-AGRGGLGQAGAGAAAAAGGAGGGYGGGLGQGGAGAGAGAGAGGGYGGGLG 260
QY 290 QGTSGRGLGQAGAGAAAAAGGAGGGYGGGLGSGQTSPPGYGP-GQOTSRRGL 348
Db 261 QG-----AGRGEGAG-----AAAAAGGAGGGYGGGLGQGG-AGQGGYGGGLGQGGAGRGL 311
QY 349 GGQAGAAAA-----AAAAAGGAGGGYGGGLGSGQTSRRGLGQGG 390
Db 312 GGQAGAAAAAGGAGGGYGGGLGQGGAGAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGQGG 370
QY 391 AGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGG 450
Db 371 AG---AVAAAAAGGAGGGYGGGLGSGQ-----AGRGGGAG-----AAAAAGGAGGGYGG 419
QY 451 LGSQGTSGRGLGQAGAGAAAAAGGAGGGYGGGLGSGQT-----SGP 497
Db 420 LQNG-AGRGGLGQGGAG-----AAAAAGGAGGGYGGGLGSGQ-----AAAAAGGAGGGYGG 475
QY 498 GGYGP-GQOTSRRGLGQGGAGAAAAAGGAGGGYGGGLGSGQTSRRGLG 550
Db 476 GGYGGGLGSGAGR---GGQAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGSGQ-AGRGGLGSGQ 531
QY 551 GGQAGAAAA-----AAAAAGGAGGGYGGGLGSGQ-----SGP 497
Db 532 GGQAGAAAAAGGAGGGYGGGLGQGGAGAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGSGQ 587
QY 591 OGAGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGG 650
Db 588 QGAG-----AAAAAGGAGGGYGGGLGSGQV-GRGLGQGGAG-----AAAAAGGAGGGYGG 636
QY 651 GGLGS 655
Db 637 GGVGS 641

RESULT 15

AAV59070
ID AAV59070 standard; Protein; 718 AA.

AC AAV59070;

DT 08-MAR-2000 (first entry)

XX N. clavipes spider silk protein 1.

DE Spider silk protein; dragline silk protein; major ampullate gland; fiber.

XX Nephila clavipes.

XX US5989894-A.

XX 23-NOV-1999.

XX 04-OCT-1994; 94US-0317844.

XX 15-APR-1991; 91US-0684819.

XX 20-APR-1990; 90US-0511792.

XX (UWY-) UNIV WYOMING.

XX Hinman MB, Xu M, Lewis RV;

XX WPI: 2000-061225/05.

XX N-PSDB; AAZ38195.

XX Isolated DNA, vector and transformed cell encoding for and useful in

XX the production of spider silk protein -

PS Claim 1; Fig 6A-D; 65pp; English.

XX The invention provides isolated cDNA molecules coding for spider silk
CC proteins. The spider silk proteins are characterized by repeating alpha
CC and beta regions and optional variable regions. The DNA sequences are
CC useful in the production of spider silk proteins by recombinant DNA
CC techniques. The recombinant spider silk proteins may be used for the
CC production of fibers. The present sequence represents the spider silk
CC protein 1, derived from the major ampullate gland of Nephila clavipes.

XX Sequence 718 AA;

Query Match 61.1%; Score 2135; DB 21; Length 718;

Best Local Similarity 68.38; Pred No. 1.5e-148;

Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QCAGAAAAAGGAGGGYGGGLGSGQT--SGRGLGQGGAG-AAAAAGGAGGGYGG 79

Db 1 QGAG-----AAAAAGGAGGGYGGGLGQGGAGGGYGGGLGQGGAGAGAGAGGAG 56

QY 80 GYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGG 139

Db 57 GYGGGLGSGQ-----AGRGGGAG-----AAAAAGGAGGGYGGGLGSGQ-AGRGGLGQGG 107

QY 140 GAAAAAGGAGGGYGGGLGSGQTSPPGYGP-GQOTSRRGLGQGGAGAGAAAAAG 199

Db 108 G-----AAAAAGGAGGGYGGGLGSGQ-----AGRGGG-----AAAA 142

QY 200 AGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGGGLGSGQTS 250

Db 143 AGGAGGGYGGGLGSGQ-AGRGGLGQGGAGAGAAAAAGGAGGGYGGGLGQGGAGGGYGG 201

QY 251 SGTSGRGLGQGGAGAGAAAA-----AAAAAGGAGGGYGGGLGSGQTSRRGLG 289

Db 202 SQG-AGRGGLGQGGAGAGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAG 260

QY 290 QGTSGRGLGQGGAGAGAAAAAGGAGGGYGGGLGSGQTSPPGYGP-GQOTSRRGL 348

Db 261 QG-----AGRGEGAG-----AAAAAGGAGGGYGGGLGSGQ-AGQGGYGGGLGSGQGG 311

QY 349 GGQAGAAAA-----AAAAAGGAGGGYGGGLGSGQTSRRGLGQGG 390

Db 312 GGQAGAAAAAGGAGGGYGGGLGQGGAGAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGSGQ 370

QY 391 AGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGG 450

Db 371 AG---AVAAAAAGGAGGGYGGGLGSGQ-----AGRGGGAG-----AAAAAGGAGGGYGG 419

QY 451 LGSQGTSGRGLGQGGAGAGAAAAAGGAGGGYGGGLGSGQT-----SGP 497

Db 420 LQNG-AGRGGLGQGGAG-----AAAAAGGAGGGYGGGLGSGQ-----AAAAAGGAGGGYGG 475

QY 498 GGYGP-GQOTSRRGLGQGGAGAAAAAGGAGGGYGGGLGSGQTSRRGLG 550

Db 476 GYGGGLGSGAGR---GGQAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGSGQ-AGRGGLGSGQ 531

QY 551 GGQAGAAAA-----AAAAAGGAGGGYGGGLGSGQTSRRGLGSG 590

Db 532 GGQAGAAAAAGGAGGGYGGGLGQGGAGAGAGAAAAAGGAGGGYGGGLGSGQ-AGRGGLGSGQ 587

QY 591 OGAGAAAAAGGAGGGYGGGLGSGQTSRRGLGQGGAGAGAAAAAGGAGGGYGG 650

Db 588 QGAG-----AAAAAGGAGGGYGGGLGSGQV-GRGLGQGGAG-----AAAAAGGAGGGYGG 636

QY 651 GGLGS 655

Db 637 GGVGS 641

Search completed: December 18, 2002, 16:17:12
Job time : 51.4697 secs

This Page Blank (uspto)

[illegible]

STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 61.1%; Score 2135; DB 1; Length 718;
Best Local Similarity 68.3%; Pred. No. 7.8e-151;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGCGGGLGSGGT--SGRGLGGGAG--AAAAAAGAGGAGG 79
DB 1 QGAG-----AAAAAGGAGGCGGGLGGGAGGCGGAGGAGGAGGAGGAGGAGG 56
QY 80 GYGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 139
DB 57 GYGGLGSGT--AGRGGAG-----AAAAAGGAGGCGGGLGSGGT--SGRGLGGG 107
QY 140 GAAAAAAGAGGAGGCGGGLGSGGTSGPGYGPQGTSGRGLGGGAGGAGGAGGAGG 199
DB 108 G-----AAAAAGGAGGCGGGLGSGT-----AGRGGAG-----AAAA 142
QY 200 AGGAGCGGGLGSGGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 250
DB 143 AGGAGCGGGLGSGT--AGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201
QY 251 SGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 289
DB 202 SQG--AGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 260
QY 290 QGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 348
DB 261 QG-----AGRGGAG-----AAAAAGGAGGCGGGLGSGT--AGRGGLGGG 311
QY 349 GGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 390
DB 312 GGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370
QY 391 AGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAG 450
DB 371 AG-----AVAAAAAGGAGGCGGGLGSGT--AGRGGAG-----AAAAAGGAG 419
QY 451 LGSQGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 497
DB 420 LGNGG--AGRGGLGGGAG-----AAAAAGGAGGCGGGLGSGT--AGRGGLGGG 475
QY 498 GYGFP--GQGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 550
DB 476 GYGGLGSGQAGR---GGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531

QY 551 GGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 590
DB 532 GGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 587
QY 591 QGAGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAG 650
DB 588 QGAG-----AAAAAGGAGGCGGGLGGGAG--GRGLGGGAG-----AAAA 636
QY 651 GGLGS 655
DB 637 GGVGS 641

RESULT 4

US-08-317-844B-2
Sequence 2, Application US/08317844B
Patent No. 5989894
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5989894th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,844B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-844B-2

Query Match 61.1%; Score 2135; DB 2; Length 718;
Best Local Similarity 68.3%; Pred. No. 7.8e-151;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGCGGGLGSGGT--SGRGLGGGAG--AAAAAAGAGGAGG 79
DB 1 QGAG-----AAAAAGGAGGCGGGLGGGAGGCGGAGGAGGAGGAGGAGGAGG 56
QY 80 GYGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139
DB 57 GYGGLGSGT--AGRGGAG-----AAAAAGGAGGCGGGLGSGGT--AGRGGLGGG 107
QY 140 GAAAAAAGAGGAGGCGGGLGSGGTSGPGYGPQGTSGRGLGGGAGGAGGAGGAGGAG 199
DB 108 G-----AAAAAGGAGGCGGGLGSGT-----AGRGGAG-----AAAA 142

RESULT 6
US-08-556-978B-21
; Sequence 21, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-21

Query Match	59.4%;	Score 2076;	DB 4;	Length 606;
Best Local Similarity	69.7%;	Pred. No. 1.5e-146;		
Matches 477;	Conservative 11;	Mismatches 74;	Indels 122;	Gaps
QY	19	GLGCGAGAAAAAAGAGAGCGGGYGGILGSGQTSRGGLGCGGACAAAAAAGAGAG 78		
Db	3	GRGGQAG---AAAAAGAGGGYGGILGSGG-AGRGLGGQAG---AAAAAAGAGAG 54		
QY	79	QGGYGGILGSGQTSRGGLGCGGACAAAAAAGAGAGCGGCGLGSGQT---SGRGLGG 136		
Db	55	Q---GGLGSGQAGAGAAAAAGAGAGCGGCGLGSGQAGGGYGGILGSGGAGGAG 100		
QY	137	QGA---GAIAAAAAAGAGAGCGGCGLGSGQT---SGPGGYGPGQTT---SGRGLG 185		
Db	101	QAGRGCGAGAAAAAAGAGAGCGGCGLGSGAGRGGLGCGGACAAAAAAGAGCGGILG 160		
QY	186	GCGAGAAAAAAGAGAGCGGCGGILGSGQT---SGRGLGAGGCA---GAIAAAAAAGAG 240		
Db	161	SGAGGCGAGAAAAAGAGAGCGGCGGILGSGAGGGYGGILGSGAGRGCGGACAAAAAGG 220		
QY	241	AGGGYGGILGSGQTSRGGLGCGGACAAAAAAGAGAGCGGCGLGSGQTSRGRLGG 300		
Db	221	AGGGYGGILGSG-ACRGLGGQAG---AAAAAAGAGAGQ-----GGLG 262		
QY	301	QAGAAAAAAGAGAGCGGCGGILGSGQTSPPGYGP---GQTSRGGLGCGGACAAAAA 359		
Db	263	QAGAGAGAAAAAGAGAGCGGCGGILGSG-AGGGYGGILGSGACGR---GGCAG---A 314		

Qy	360	AAAAAGAGAGGGYGGGLGSQGT--SRRGGLGGGAGAGAAAAAAGGAGGGYGGGLGSQGT	419
Db	315	AAAAAGAGAGGGYGGGLGSQ--AGRGGLGGGAG--AAAAAAGGAGQ-----	358
Qy	420	GRGGLGGGAGAGAAAAAAGAGAGGGYGGGLGSQGT--SRRGGLGGGAG--GAAAAA	474
Db	359	--GGLGGGAGAGAGAAAAAAGAGAGGGYGGGLGSQAGGGYGGGLGGGAGGGGAGAA	416
Qy	475	AAAAAGAGAGGGYGGGLGSQGT--SGPGGGYGGPQQT-----SRRGGLGGGAGAGAAAAA	526
Db	417	AAAAAGAGAGGGYGGGLGSQAGRGGLGGGAGAGAAAAAAGAGAGGGYGGGLGGGAGAGAA	476
Qy	527	AAGAGAGGGYGGGLGSQGT--SRRGGLGGGAG--GAAAAAAGAGAGGGYGGGLGSQ	581
Db	477	AAGAGAGGGYGGGLGSQAGGGYGGGLGSQAGRGGGGAGAGAAAAAAGAGGGYGGGLGSQ	536
Qy	582	TSGRGLGGGAGAGAAAAAAGAGAGGGYGGGLGSQGT--SRRGGLGGGAGAGAAAAA	641
Db	537	--AGRGGLGGGAG--AAAAAAGGAGQ-----GGLGGGAGAGAGAGAA	578
Qy	642	AGGAGGGYGGGLGSQGTSGPGGYG	665
Db	579	AGGAGGGYGGGLGSQ--AGQGGYG	601
RESULT 7			
US-09-247-806-4			
; Sequence 4, Application US/09247806			
; Patent No. 6280747			
; GENERAL INFORMATION:			
; APPLICANT: PHILLIPPE, Michel			
; APPLICANT: GARSON, Jean-Claude			
; APPLICANT: ARRAUDEAU, Jean-Pierre			
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT			
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN			
; TITLE OF INVENTION: ANALOG			
; FILE REFERENCE: 6388-0365-0			
; CURRENT APPLICATION NUMBER: US/09/247,806			
; CURRENT FILING DATE: 1999-02-11			
; EARLIER APPLICATION NUMBER: FR 98/01614			
; EARLIER FILING DATE: 1998-02-11			
; NUMBER OF SEQ ID NOS: 14			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 606			
; TYPE: PRT			
; ORGANISM: Nephila clavipes			
; US-09-247-806-4			

Query Match	59.4%	Score	2076;	DB 4;	Length	606;			
Best Local Similarity	69.7%	Pred.	No. 1.5e-146;						
Matches	477;	Conservative	11;	Mismatches	74;	Indels	122;	Gaps	30;

Qy	19	GLGQGAGAAAAAAGAGAGCGGCGGLGSGQGTSGRGGLGGCGGAGAAAAAAGAG	78
Db	3	GRGQGGAG---AAAAAGAGAGCGGCGGLGSGQ-AGRGGLGGCGAG---AAAAAAGAGAG	54
Qy	79	QGGYGGLGSGQTSRGRGLGGGAGAAAAAAGAGAGCGGCGGLGSGQGT--SGRGGLGG	136
Db	55	Q-----GGLGSGQGGAGAAAAAAGAGAGCGGCGGLGSGQGGCGGCGG	100
Qy	137	QGA---GAAAAAAGAGAGCGGCGGLGSGQGT--SGPGYGPGCQGT-----SGRGGLG	185
Db	101	QAGRGGGAGAAAAAAGAGAGCGGCGGLGSGGACRGGLGGCGAGAAAAAAGAGAGCGGLG	160
Qy	186	GQGAGAAAAAAGAGAGCGGCGGLGSGQGT--SGRGGLGGQGA---GAAAAAAGAG	240
Db	161	SGAGQGAGAAAAAGAGAGCGGCGGLGSGQAGCGGCGGLGSGGACRGCGGAGAAAAAGG	220
Qy	241	AGCGGCGGLGSGQTSRGRGLGGGAGAAAAAAGAGAGCGGCGGLGSGQGTSGRGGLGG	300
Db	221	AGCGGCGGLGSGQ-AGRGGLGGCGAG---AAAAAAGAGAG-----GGGLGS	262

Query Match		56.5%;	Score 1972.5;	DB 4;	Length 606;
Best Local Similarity		64.2%;	Pred. No. 6.7e-139;		
Matches		462;	Conservative	8;	Mismatches 35; Indels 215; Gaps
Qy	38	GACGGGYGGLGSGQTS	SRGGGLGGQAGAGAAAAA	AAAAAGAGCGGCGYGG	LGSGQTSGRGGLG 97
Db	1	GAGGGGYGGLGSGQ-	AGRGGLGGQAG--	AAAAAAGGAGQ-	-----GGLG 42
Qy	98	GQAGAAAAA	AAAAAGAGCGGCGYGG	LGSGQTS	SRGGGLGGQAGAAAAA
Db	43	SQAGAGAGAAAAA	AGAGCGGCGYGG	LGSGQ-	----AGRGCGAG--
Qy	158	YGLGSGQTS	SGPGYGP-QQ	TS	SRGGGLGGQAGAAAA
Db	95	YGLGSGQ-	AGQGGYGG	LGSGQAGRGGLGGQAGAAAAA	AGAGCGGCGYGG
Qy	196	AAAAAGAGCGGCGYGG	LGSGQTS	SRGGGLGGQAGAAAAA	AAAAAGAGCGGCGYGG
Db	154	AAAAAGAGCGGCGYGG	LGSGQ-	----AGRGCGAG--	AAAAAAGAGCGGCGYGG
Qy	255	-----	SRGGGLGGQAGAAAAA	AAAAAGAGCGGCGYGG	LGSGQTS
Db	206	QGGYGG	LGSGQAGRGGLGGQAG-	AAAAAAGAGAGQ-	-----GGLG
Qy	304	GAAAAA	AAAAAGAGCGGCGYGG	LGSGQTS	SGPGYGP
Db	249	GQAGAAAAA	AGAGCGGCGYGG	LGSGQ-	-----AGRGCGAG--
Qy	364	AGGAGCGGCGYGG	LGSGQT-	-----	SRGGGLGGQAGAAAAA
Db	289	AGGAGCGGCGYGG	LGSGQAGCGGCGYGG	LGSGQAGRGGLGGQAG--	AAAAAAGAGAGQ-
Qy	412	GLGSGQTS	SRGGGLGGQAGAAAAA	AAAAAGAGCGGCGYGG	LGSGQTS
Db	342	-----	GGLGSGQAGCGAGAAAAA	AGAGCGGCGYGG	LGSGQ-
Qy	472	AAAAAAGAGCGGCGYGG	LGSGQTS	SGPGYGP-QQ	TS
Db	385	AAAAAAGAGCGGCGYGG	LGSGQ-	AGQGGYGG	LGSGQAGRGGLGGQAGAAAAA
Qy	523	-----	AAAAAAGAGCGGCGYGG	LGSGQTS	SRGGGLGGQAGAAAAA
Db	443	GGLGSGQAGCGAGAAAAA	AGAGCGGCGYGG	LGSGQ-	----AGRGCGAG--
Qy	570	GQGGYGG	LGSGQT-	-----	SRGGGLGGQAGAAAA
Db	495	GQGGYGG	LGSGQAGCGGCGYGG	LGSGQAGRGGLGGQAGAAAAA	AGAGCGGCGYGG
Qy	599	--	AAAAAAGAGCGGCGYGG	LGSGQTS	SRGGGLGGQAGAAAAA
Db	555	AGAAAAA	AGAGCGGCGYGG	LGSGQ-	AGRGCGAG--

RESULT 8

yy 304 AGGAGGGGTGGTGGGT-----SGRGGGGCGAGAAAAAAGCAGGGYG 411
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

```

Db      289 AGGAGGGYGGGLGSQAGGGYGGGLGSQGAGRGGLGSQAG---AAAAAAGGAG- --- 341
Qy      412 GLGSGQTSGRGGLCGCAGAAAAAAAGACAGCGGYGGGLGSQGTSGRGGLGGCAGAAA 471
Db      342 -----GGLSQAGCAGAAAAAAGGAGCGGYGGGLGSQ-----AGRGGCAG-- 384
Qy      472 AAAAAAAGGAGCGGYGGLSQGTSGPGGYGP-CQDTSGRGGLGGCAGAAAA----- 522
Db      385 -AAAAAAGGAGCGGYGGLSQG-AGCGGYGGGLGSQAGRGGLGGCAGAAAAAAGGAGC 442
Qy      523 -----AAAAAAGGAGCGGYGGGLGSQGTSGRGGLGSQAGAAAAAAGA 569
Db      443 GGLGSQCAGCAGAAAAAAGACGGYGGLGSQ-----AGRGGCAG----AAAAAAGA 494
Qy      570 GGCGYGGGLGSQGT-----SGRGGLGGCAGAAA----- 598
Db      495 GGCGYGGLSQAGCGGYGGLSQAGRGGLGSQAGAAAAAAGGAGCGGLGSQAGCG 554
Qy      599 --AAAAAAGGAGCGGYGGLSQGTSGRGGLGGCAGAAAAAAGGAGCGGYGGLSQ 656
Db      555 AGAAAAAAGGAGCGGYGGLSQG-----AGRGGCAG-----AAAAAAGGAGCGGYGGLSQ 606

RESULT 9
US-09-247-806-8
; Sequence 8, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 606

```


Qy 136 GGGAGAAAAAAGAGAGGGGGLGSGQTSRGGTSGRGGGAGGAGAAA 195
Db 184 GYGRGAGAGAAA--GAGAGGYGGGGYGAGAGAGAAAAAGAGSGGAGTGRGAGAGA 241
Qy 196 AAAAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAAA--AAAAAG 239
Db 242 GAAAGAGAGAGSYG-----GQYGAGAGAGAAAAAAXXXXXXXXXXGAGAGAG 293
Qy 240 GAGGGYG-----GLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGYG 286
Db 294 YGGGGYGAGAGAGAAAAAGAGAGAGGGTGRGAGAGAAAAAGAGAGAGAGAGAGAG 353
Qy 287 ---LGSQTSRGLG---CGAGAAAAAAGAGAGAGGGGGLGSGQTSRGGTSGPGGYPG 339
Db 354 GAAAGAGAGAGAGGGTGRGAGAGAGAAAAAGAGAGAGGGGAGGGYG--AGAGAGAAA 412
Qy 340 QOTSGRGLGGGAGAAAAAAGAGAGGGGGLGSGQ-----TSRGGGLGG 388
Db 413 AAGAGAGAGGGTGRGAGAGAGAAA--GAGAGGYGGGGYGAGAGAGAAAAATGAGGAGG 470
Qy 389 QGAGAAAAAAGAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGYG 448
Db 471 YGRGAGAGAGAGAGAGAGTGG-----AGYGGGGGYGAGAGAGAGGGGAGGGAGY 521
Qy 449 G-GLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG---LGSQTSRGG- 503
Db 522 GRGAG-----AGAGAAAGAGAGAGAGAGAGGGGAGAGAGAGAGAGAGAGAGAG 577
Qy 504 QOTSGRGLGGGAGAAAAAAGAGAGGGGGLGSGQ-----GLGS-----QGTSGRGGGLG-GQG 554
Db 578 YSRGRAGAGAGAGAGAGAGAGAGGGGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
Qy 555 AGAAAAAAGAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGYG 614
Db 638 AGAAAGAGAGAGAGAGAGY-----GGGGYGAGAGAGAAAAAAGAGAGAGAGAGAG 683
Qy 615 LGSQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG 651
Db 684 -RGAGAGGYGGGGYGAGAGAGAGAAAA--GAGAGGYG 717

RESULT 13

US-08-458-298-2
; Sequence 2, Application US/08458298
; Patent No. 5756677
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,298
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,747
; FILING DATE: 14-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.

; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: N. clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..309
; US-08-458-298-2

Query Match 40.2%; Score 1405; DB 1; Length 832;
Best Local Similarity 50.1%; Pred. No. 6.7e-97;
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;
Qy 21 GQGGAGAAAAAAGAGAGGGGGLGSGQTSRGGTSGRGGGAGAGAAAAAAGAGAGGG 80
Db 79 GYGAG--AGAVAAAGAGAGGYG---RGAGGYGGGGYGAGAGAGAAAA--GAGAG 129
Qy 81 GYGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG-----GLGSGQTSRGG 135
Db 130 GAGGY-----GRGAGAGAGAAAGAGAGAGGGGAGGGYGAGAGAGAAAAAGAGAGG 183
Qy 136 GQGGAGAAAAAAGAGAGGGGGLGSGQTSRGGTSGPGGYPGQOTSGRGGGLGGGAGAAAA 195
Db 184 GYGRGAGAGAGAAA--GAGAGGYGGGGYGAGAGAGAAAAAAGAGAGGGGAGTGRGAGAGA 241
Qy 196 AAAAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAA--AAAAAG 239
Db 242 GAAAGAGAGAGSYG-----GQYGAGAGAGAAAAAAXXXXXXXXXXGAGAGAGAG 293
Qy 240 GAGGGYG-----GLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG----- 286
Db 294 YGGGGYGAGAGAGAAAAAGAGAGAGGGTGRGAGAGAGAGAGAGAGAGGGGAGGGYGAGAGA 353
Qy 287 ---LGSQTSRGGGLG---GQGGAGAAAAAAGAGAGGGGGLGSGQTSRGGTSGPGGYPG 339
Db 354 GAAAGAGAGAGAGGGTGRGAG 412
Qy 340 QOTSGRGLGGGAGAAAAAAGAGAGGGGGLGSGQ-----TSRGGGLGG 388
Db 413 AAGAGAGAGGGTGRGAGAGAGAGAAA--GAGAGGYGGGGYGAGAGAGAGAAAAATGAGGAGG 470
Qy 389 QGAGAAAAAAGAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGYG 448
Db 471 YGRGAGAGAGAGAGAGAGTGG-----AGYGGGGGYGAGAGAGAGGGGAGGGAGY 521
Qy 449 G-GLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG---LGSQTSRGG- 503
Db 522 GRGAG-----AGAGAAAGAGAGAGAGAGAGGGGAGAGAGAGAGAGAGAGAGAGAG 577
Qy 504 QOTSGRGLGGGAGAAAAAAGAGAGGGGGLGSGQ-----GLGS-----QGTSGRGGGLG-GQG 554
Db 578 YSRGRAGAGAGAGAGAGAGAGAGGGGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
Qy 555 AGAAAAAAGAGAGAGGGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGGYG 614
Db 638 AGAAAGAGAGAGAGAGY-----GGGGYGAGAGAGAAAAAAGAGAGAGAGAGAGAG 683
Qy 615 LGSQTSRGGGLGGGAGAGAAAAAAGAGAGGGGAGGGYG 651
Db 684 -RGAGAGGYGGGGYGAGAGAGAGAAAA--GAGAGGYG 717

Db 193 -----GPSGPGSAAAAAAGPGGYGPGQGPGGY- PGQGPGGYGPQGPSPGSA 246
QY 235 AAAAGGAGGGYGLGSGGTSGRGLG- GQAGAGAAAAAAGAGGAGGGYGLGSGQTS 293
Db 247 AAAAAAGPGQGPGGY- PGQGPGGYGPQGPSPGSA AAAAAAGPGGYG- GQGP 304
QY 294 GRGGLGGGAGAAAAAAGGAGGGYGLGSGGTSGRGLG 353
Db 305 GYGP- GQGPSPGSA AAAAAAGPGGYG- PGQGPGGYGPQGPSPGSA 359
QY 354 GAAAAAAGGAGGGYGLGSGGTSGRGLG- GQAGAGAAAAAAGAGGAGGGYGLG 412
Db 360 SGPGSAAAAA- GPQGPGGY- PGQGPGGYGPQGPSPGSA AAAAAAGPGGYG 417
QY 413 LGSQTSRGLGSGGAGAAAAAAGGAGGGYGLGSGGTSGR- GLGGQAG 468
Db 418 -GQGPGGYGP- GQGPSPGSA AAAAAAGPGGYG- GQGPGGYGPQGPSPGSA 474
QY 469 -----AAAAAAGGAGGGY- GLGSGTS-----GP 500
Db 475 QGPSPGSA AAAAAAGPGQGPGGYGPQGPSPGSA AAAAAAGPGY 534
QY 501 GPQQTSGRGLGSGGAGAAAAAAGGAGGGYGLGSGGTSGR- GLGGQAG 556
Db 535 GPGQGPGGYGPQGPSPGSA AAAAAAGPGGYG- GQGPGGYGPQGPSPGSA 593
QY 557 -----AAAAAAGGAGGGYGLGSGGTSGRGLG- GQAGAGAAAAAAGAGGAGGG 611
Db 594 QGPSPGSA AAAAAAGPGQGPGGY- PGQGPGGYGPQGPSPGSA AAAAAAGPG 652
QY 612 YG-GLGSGTSGRGLGSGGAGAAAAAAG-----GAGGGYGLGSGGTSGPGGYG 666
Db 653 YGPQGPGGYGPQGPSPGSA AAAAAAGPGGYGPQGPGGY- PGQGPGGYGP 711
QY 667 GQ 669
Db 712 GQ 714

Search completed: December 18, 2002, 16:23:06
Job time : 18.672 secs

This Page Blank (uspto)

Db 514 -----GSAGYGFQSAS 525

RESULT 7

US-10-052-586-97

; Sequence 97, Application US/10052586

; Patent No. US20020127584A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3430R1C1

; CURRENT APPLICATION NUMBER: US/10/052,586

; CURRENT FILING DATE: 2002-01-15

; PRIOR APPLICATION NUMBER: 60/039263

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/059266

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/062250

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 60/063120

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063121

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063486

; PRIOR FILING DATE: 1997-10-21

; PRIOR APPLICATION NUMBER: 60/063540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063541

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063544

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063564

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063734

; PRIOR FILING DATE: 1997-10-29

; PRIOR APPLICATION NUMBER: 60/063870

; PRIOR FILING DATE: 1997-10-31

; PRIOR APPLICATION NUMBER: 60/064103

; PRIOR FILING DATE: 1997-10-31

; PRIOR APPLICATION NUMBER: 60/065311

; PRIOR FILING DATE: 1997-11-13

; PRIOR APPLICATION NUMBER: 60/066120

; PRIOR FILING DATE: 1997-11-21

; PRIOR APPLICATION NUMBER: 60/066466

; PRIOR FILING DATE: 1997-11-24

; PRIOR APPLICATION NUMBER: 60/066772

; PRIOR FILING DATE: 1997-11-24

; PRIOR APPLICATION NUMBER: 60/069335

; PRIOR FILING DATE: 1997-12-11

; PRIOR APPLICATION NUMBER: 60/069425

; PRIOR FILING DATE: 1997-12-12

; PRIOR APPLICATION NUMBER: 60/069870

; PRIOR FILING DATE: 1997-12-17

; PRIOR APPLICATION NUMBER: 60/068017

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/077450

; PRIOR FILING DATE: 1998-03-10

; PRIOR APPLICATION NUMBER: 60/077632

; PRIOR FILING DATE: 1998-03-11

; PRIOR APPLICATION NUMBER: 60/077649

; PRIOR FILING DATE: 1998-03-11

; PRIOR APPLICATION NUMBER: 60/078886

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 60/078939

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 60/079664

; PRIOR FILING DATE: 1998-03-27

; PRIOR APPLICATION NUMBER: 60/079786

; PRIOR FILING DATE: 1998-03-27

; PRIOR APPLICATION NUMBER: 60/080107

; PRIOR FILING DATE: 1998-03-31

; PRIOR APPLICATION NUMBER: 60/080194

; PRIOR FILING DATE: 1998-03-31

; PRIOR APPLICATION NUMBER: 60/080327

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/080333

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/081049

; PRIOR FILING DATE: 1998-04-08

; PRIOR APPLICATION NUMBER: 60/081070

; PRIOR FILING DATE: 1998-04-08

; PRIOR APPLICATION NUMBER: 60/081195

; PRIOR FILING DATE: 1998-04-09

; PRIOR APPLICATION NUMBER: 60/081838

; PRIOR FILING DATE: 1998-04-15

; PRIOR APPLICATION NUMBER: 60/082568

; PRIOR FILING DATE: 1998-04-21

; PRIOR APPLICATION NUMBER: 60/082569

; PRIOR FILING DATE: 1998-04-21

; PRIOR APPLICATION NUMBER: 60/082704

; PRIOR FILING DATE: 1998-04-22

; PRIOR APPLICATION NUMBER: 60/082797

; PRIOR FILING DATE: 1998-04-22

; PRIOR APPLICATION NUMBER: 60/083322

; PRIOR FILING DATE: 1998-04-28

; PRIOR APPLICATION NUMBER: 60/083495

; PRIOR FILING DATE: 1998-04-29

; PRIOR APPLICATION NUMBER: 60/083496

; PRIOR FILING DATE: 1998-04-29

; PRIOR APPLICATION NUMBER: 60/083499

; PRIOR FILING DATE: 1998-04-29

; PRIOR APPLICATION NUMBER: 60/083559

; PRIOR FILING DATE: 1998-04-29

; PRIOR APPLICATION NUMBER: 60/084366

; PRIOR FILING DATE: 1998-05-05

; PRIOR APPLICATION NUMBER: 60/084414

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: 60/084639

; PRIOR FILING DATE: 1998-05-07

; PRIOR APPLICATION NUMBER: 60/084640

; PRIOR FILING DATE: 1998-05-07

; PRIOR APPLICATION NUMBER: 60/084643

; PRIOR FILING DATE: 1998-05-07

; PRIOR APPLICATION NUMBER: 60/085573

; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/085579

; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/085580

; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/085582

; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/085700

; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/086023

; PRIOR FILING DATE: 1998-05-18

; PRIOR APPLICATION NUMBER: 60/086392

; PRIOR FILING DATE: 1998-05-22

; PRIOR APPLICATION NUMBER: 60/086486

; PRIOR FILING DATE: 1998-05-22

; PRIOR APPLICATION NUMBER: 60/087098

; PRIOR FILING DATE: 1998-05-28

; PRIOR APPLICATION NUMBER: 60/087208

; PRIOR FILING DATE: 1998-05-28

; PRIOR APPLICATION NUMBER: 60/087609

;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/078886
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078939
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079664
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079786
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/080107
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080194
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080327
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080333
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/081049
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081070
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081195
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081838
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082568
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082569
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082704
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082797
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086023
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/086392
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086486
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087098
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087208
;; PRIOR FILING DATE: 1998-05-28

;; PRIOR APPLICATION NUMBER: 60/087609
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 19.6%; Score 686; DB 12; Length 1300;

Best Local Similarity 28.1%; Pred. No. 5.7e-33;

Matches 277; Conservative 16; Mismatches 347; Indels 346; Gaps 38;

QY 14 ASGRGGLGGGAGAAAAAAGAGAG-----QGGYGLGSGGTSGRG-----GLGGGGA 63

DB 182 ACGAGACGGAGGAGGTCCCAAGATGCCCTCCGACCACTAAAGAGAGAAATCGTGGGA 241

QY 64 GAAAAAAGAGAGGAGGGYGGGLGSGQ-----TSRRGG-LGCGGAGAAAAAA 109

DB 242 ATAAGAACTTCCACGAGGTGATGCTGCTCACACTGTAGAAACCTGTGTCAAGAACT 301

Qy 110 AAAGAGGGYGLGSGQTSRGGGLGGGAGAGAAAAAGGAGGGYGLGSGQTSR 169
Db 302 GGGGACCGCTTCCAGCTGCTGGTGGCCAGCAGGCTTCTGGAGAGTGTGCTGGA 361
Qy 170 GG-----YGPQQT----- 178
Db 362 GGACCATCTGCCAAGAACCAACCCACCACCATCGTGCATGACAAAGTGTCAACCTCA 421
Qy 179 -----SRGGGLGGQAG-----AAAAAAGGAGGQ----- 206
Db 422 TCCAGTCTCGGCTGACGCGTCCGAGCTGCCGCTGACAGGTGTGCTCACCCTCT 481
Qy 207 -GYGGLGSGQTSRGG-----LGOGAGAAAAAAGGAGGGYGLGSGGT--- 254
Db 482 ATGAGACCTCGGAGGAAAGCGCTGGAGTTCGCCATGACTGACCTGGACATGTGTGAC 541
Qy 255 -----SRGG-LGGQ-----AGA-AAAAAAGGAGGQGY----- 284
Db 542 CCATCCACACACCCAGAGGACCGTGTCAACTCAGAGACACAATCAGGACGATTCGT 601
Qy 285 -GGLGSGQ-----TSRGG----- 297
Db 602 GGGCACTGACTCCAGCCAGCAAGAGGACTCTGGCCAGCATCTGCCCTCTGCCGCC 661
Qy 298 -----LGQOG-----AGAAAAAAGGAGGAG-OGGYGLGSGQTSR 334
Db 662 CCCATCTCTCCGCTGACAGCCCATAGCACCACCCGCAACAGATGGGAAGCTCG 721
Qy 335 GYGFQOQ-TSRGGGLGGGAGAAAAAAGGAGGGYGLGSGQTSR-RGGLGGQ-- 390
Db 722 CAGTGAGCTGAGATGTGAGTGGAGCTGAGGCTGATGCGGAGTGTGACGAGCT 781
Qy 391 AGAAAAAAGG-----AGOGYGLGSGQTSRGGGLGGGAGAAAAAAGG 440
Db 782 GGTGCCACCCAGGCGGAGCCGAGACCTGGAGTGTGCGAG-----GAGCTCAACCCGA 837
Qy 441 GGAGGGYGLGSGQTSRGG--LGOGAGAG-----AAAAAAGG 478
Db 838 CTGCCAGGACATGACAGCGGCTCTGAGTATACCTCTCGGGCCCCATGCCCCAA 897
Qy 479 GGAG--OGGYGLG-----SQTSRGGYGPQQTSGRGGGLG-----OCAGAAA 521
Db 898 GGAGCCCTTACAGCCACACTGCCAGTCGAGGCTTGCTGGAGGCTGCCCCACAGTGA 957
Qy 522 AAAAAAGGAG-----OGGYG-----GLGSGQTSRGG 549
Db 958 ATTCTGCCGAGCCCTATTGCTCCTACCTGCTCTGCTGATGGGCCCATGGCTTG-GC 1016
Qy 550 LGG-----OGAGAAAAAAGGAGGGYGLG-----QTSRGGGLGGGAGAAAAA 600
Db 1017 TGGCCACTGAGGTAAGGTGTGGAGGTGTGGAGGCCCTCGAGAGCTGCGCGGCCAG 1076
Qy 601 AAAAGGAGGGY-----GLGSGQTSRGGGLGGGAGAAAAAAGG----- 641
Db 1077 GTACGAAGCTCAACTCTGCCGAGAG-----GGCAGATCTCATGACGCCCGAGCTGC 1131
Qy 642 AGGAGGG-----YGLGSGQTSRGG 663
Db 1132 AGGTGAGGCTTCAGGGGATGCTGGG 1157

RESULT 10
US-09-060-854B-2
; Sequence 2, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060,854B

; CURRENT FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: B. amyloliquefaciens
US-09-060-854B-2

Query Match 19.3%; Score 675.5; DB 10; Length 1497;
Best Local Similarity 30.7%; Pred. No. 2.5e-32;
Matches 265; Conservative 22; Mismatches 330; Indels 245; Gaps 28;

Qy 25 AGAAAAAAGGAGGGYGLGSGQTSRGGGLGGGAGAAAAAAGGAG----- 77
Db 70 AATGAAAAAAGGAGAG-----GATAAGAGCTGAGAGGCAAAAAAAGTATGATCAGTTT 124
Qy 78 -----GQGYGGLGSGQ-----GTSG-----RGSL 96
Db 125 GCTGTTTCTTAGCGTTAATCTTTACGATGGCGTTCGGCAGCAGCATCTCTGCCAGGC 184
Qy 97 GQAGAGAAAAAAGGAGGGYGLGSGQTSRGGGLGGGAGAAAAAAGGAG-- 154
Db 185 GCGAGGAAATCAACCGGGAAA-AGAAATATATTCTCGGTTTAAACAGACATGAGCA 243
Qy 155 QGYG-GLGSGQTSRGGYGPQQTSG-----RGGLGGGAGAAAAAAGGA 203
Db 244 CGATGAGCGCCGCTAAGAGAAAGATGTCTTCTGAAAAAGCGGGAAGTCAAAAGC 303
Qy 204 GQGYGGLGSGQTSRGGGLG--QGAGAAAAAAGGAGGGYGLGSGQTSRGG 261
Db 304 AATCAATATGTAGACCCAGCTTCAGTCACATTAACGAA-----AAGCTGA--AAG 357
Qy 262 GQAGAAAAAAGGAGGGYGLGSGQTSRGGGLGGGAGAAAAAAGGAGGG 321
Db 358 AATTGAAAAAGACCGGAGCGTCTAGCTTGAAGAAG---ATCACGTAGCATGCT 414
Qy 322 YGLGSGQTSRGGYGPQQTSGRGGGLGGGAGAAAAAAGGAG----- 362
Db 415 ACGGCAGCTC---CGTGCTTACGGCTATACAAATTAAGAGCCCTCTCTGCATCT 470
Qy 363 -AAGG-----AGQGYGGLGSGQTSRGGGLGGGAGAAAAAAGGAG----- 406
Db 471 CAGGCTACATGATCAATGTTAAGTAGCGGTTATCGACAGCGGTATCATCTCTCT 530
Qy 407 -----QGYGGLGSGQTSRGGGLGGG----- 428
Db 531 CATCTGATTTAAGGTAGCAAGCGGAGCCAGCATGGTTCTCTGAAACAAATCTTTC 590
Qy 429 -AGAAAAAAG-----AAGGAG-----QGYGGLGSGQTSRGGGLGGGAGAAAAA 475
Db 591 CAAGACAACAACCTCTCAGGAACTCACGTTGCCGACAGTTTGGCGCTCTTAATACTCA 650
Qy 476 AAGGAG-----QGYG-----GLGSGQTSRGG-----YGPQQT 506
Db 651 ATCGTGTATTAGCGTTGCCCAAGCCAGCATCTTACGCTGTAAAGTTCTCGCTCT 710
Qy 507 SRRGLGGGAGAAAAAAGGAGGGYGLGSGQTSRGGGLGGGAG-----GAAAAAAG 564
Db 711 GACGGTCCGGCCATACAGCTGGATCATTAACGAATCGAGTGGCGCATCAACAAT 770
Qy 565 AAGG-----GQGYG-----LGSGQTSRGGGLGGG 592
Db 771 ATGACGTTATTAAATGAGCTTCGGCGGAGCTTCTGTTCTGCTCTTAAAGCGGCA 830
Qy 593 AGAAAAAAGGAG-----OGGYGGLGSGQTSRGGGLGGGAGAAAAAAGG----- 644
Db 831 GTTGATAAGCCGTGATCGGCTCTAGTCTGTTGGGAGCGCGGTAAACGAGCACT 890
Qy 645 -----AGOGYGLGSGQTSRGG 660
Db 891 TCCGCGAGCTCAAGCAGATGG 912

RESULT 11
US-10-029-217A-24
; Sequence 24, Application US/10029217A
; Patent No. US20020164735A1
; GENERAL INFORMATION:
; APPLICANT: OLSON, ERIC N.
; APPLICANT: WANG, DA-ZHI
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO A CARDIAC-SPECIFIC
; FILE REFERENCE: UTSID:695US
; CURRENT APPLICATION NUMBER: US/10/029,217A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/257,761
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-029-217A-24

Query Match 19.2%; Score 672; DB 9; Length 3907;
Best Local Similarity 30.9%; Pred. No. 7.7e-32;
Matches 258; Conservative 15; Mismatches 342; Indels 220; Gaps 33;

QY 24 GAGAAAAAAGGAGGQGG---YGGIGSQGTSRGGG-----LGGQAG-----64
DB 886 GCCAAGTCAGCAGCGGAGCCCTGG-GAAGCAGCGGAGCCCCCAGTAGCGAGCCTCT 944
QY 65 --AAAAAAGAGGAG-----QGGY-----GGLSQGTSRGGIGGGGAGAAAA 106
DB 945 CCACCTACCAATAGCAGCTCCAGCTCGGCGCCCTCGGCGCTGTGG-GCTGGCAGCTCAG 1003
QY 107 AA-----AAAAAGAGGGYGLGSGQTSRGGIGGGGAGAAAAAAGAGGAG 154
DB 1004 AACAGCACCTACTGACTGCGCAAGCCG--GGAGCCCTCGCGGCCAACCTGGAGCAGACATGA 1061
QY 155 QGGYGGIGSQGTSRGGYGPQQTSRGGIGGGGAG-----GAAAAAAGAA 199
DB 1062 AGGTGGCAGAGCTGAGCAGGAGCTGAAGTTGGATCCTGCTGCTCGGGCACCACAAA 1121
QY 200 AGAGAGGGY---GGLSQGTSRGGIGGGGAGAAAAAAGAGGAGG---YGGIGSQG 253
DB 1122 CTGAGCTGAFTGAGCGCCTTCGAGCCTATCAAGACCAAAATCAGCCCTGTGCCAGGAGCC 1181
QY 254 TSGRGG---LGGQAGAG-----AAAAAAGAGGAGGGYGGIGSQGTSG----294
DB 1182 CCAAGGCCCTGCGGCACCTCTATCTGACAAAGGCTGGCAGGTTGGT--GTAGCCTT 1239
QY 295 --RGLGG-----QAGAAAAAAGAGGAGGGYGGIGLS-----QGTSGP 333
DB 1240 CCCAGCGCCGCTGAGCAGCGGGCCGCGCTGTGGCAGCAGGCTGCTCCAGCTGA 1299
QY 334 GYGPGQQTSGRGGIGGGGAGAAAAAAGAGGAGGGYGGIGSQGTSRGGIGGGAG--392
DB 1300 GG-----TGGTGTGGCCACGCTGGCCAGCAGTGGGCTGTTGAAGTT-TGGCAGCAGCG 1352
QY 393 -----AAAAAAGAGGAGG---GYGGLSQGTSRGGIGGGG 428
DB 1353 GCTCCAGCCCCCGTGTCTCCACCCCTCGGAGGCGCTCACTGCTCAGCAGCGGCGATG 1412
QY 429 AGAA-----AAAAAAGAGGAGGGYGGIGSQGTSG-----RGLGGQAGAA 470
DB 1413 AANACTCCACCCCGGGAGACCTTTGGTGAGATGTTGACATCACCTCTGACCCAGCTGA 1472
QY 471 AAAAAAAGAGGQ-----GYGG-----LGSQGTSPGGYGPQQTSRGGIGGGQ-----516
DB 1473 CCCTGAGCGCTCGCCACATGCAGATCCTCGTGAAGAGGAGGAGGCCCCCGGGCGGCTCCT 1532
QY 517 --AGAAAAAAGAGGAGGGYGGIGSQG--TSRGGIGGGGAGAAAAAAGAGGAGGQGG 573

DB 1533 GTTCCTGAGCCCTGGGGCGCGGAGCTAGAGG-GGCGCGACAAGACCATGCTG 1591
QY 574 YGGLGSQGTSRGGIGGGGAGAAAAAAGGA--GOGGYGGLG-----SQGTSRGG 625
DB 1592 CAG-----GAGAAAGACAAGCAGATCGAGGCGCTGACGCGCATGCTCGG 1636
QY 626 LGGQAGAAAAAAGAGGAGGQGG-----YGGIGSQGTSRGGYGP 667
DB 1637 CAGAAGCAGCAGCTGGTGGAGCGCTCAAGCTGCGAGCTGGAGCAGGAGAACGAG 1691

RESULT 12
US-10-096-961-1
; Sequence 1, Application US/10096961
; Patent No. US20020155572A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; FILE REFERENCE: CL000849DIV
; CURRENT APPLICATION NUMBER: US/10/096,961
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/232,632
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/738,884
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-096-961-1

Query Match 18.8%; Score 656.5; DB 9; Length 2211;
Best Local Similarity 29.2%; Pred. No. 3.9e-31;
Matches 274; Conservative 17; Mismatches 337; Indels 309; Gaps 37;

QY 19 GLGGGAGAAAA---AAAAAG-----GAGQG-----GYGGLGSQGTSGRG 56
DB 525 GGAGTGTGACCACTCCCAACAACGACCCTGTAGAGGGGCTGAGATCGAGGAGTTCTCGC 584
QY 57 GLGGGAGAAAAAAGAGGAG--QGGYGGLS-----QGTSGRGGIGGGGAGAAAA 106
DB 585 GCGGCTCTCAAGCGCGCGGAGCTGGAGGAGATCTTCCATCAGTACTCG--GGCGAGACC 643
QY 107 AAAAAAGGAGGQ-----GYGGLGSQGTSRGGIGGGGAGAAAAAAGAGGGYGG 161
DB 644 CGCTGTGAGTGCCTCGAGCTGCTGGAGTTCTTGGAGGACCAGGGCGAGAG---GGC 699
QY 162 GSQGTSGPGG--YGPQQTSRGGG-----GGGAGAAAAA 195
DB 700 GCCACACTGGCCCGCCCGCCAGCAGCTCATTCAGACCTATGAGCTCAACGAGACAGCCAAG 759
QY 196 AAAAAAGGAG-----QGGYGG-----LGSQGTSG--RGLGGQGA--227
DB 760 CACATGAGCTGATGACACTGGATGGCTTCATGATGATGCTGTGTCGCCGAGGGGACT 819
QY 228 -----GAAAAAAGAGGAGGGYGGIGSQGTSRGGIGGGGAGAAAA-----269
DB 820 GCCTTGGACAACACCCACACGT---GTGTGTTCCAGGACATGAACACGCCCTTGCCCA 875
QY 270 -----AAAAA-----AGAGOGGYGGIGLSQG 291
DB 876 CTACTTCTCTCTCTCCCAACAACACCTATCTGACTGACTCCACAGATCGGGGGGCCAG 935
QY 292 TSGRG--GLGG-----QGAGAAAAAAGAGGAGG-----GYGGLGSQGTSGPG 334
DB 936 CACACCGAGGCTATGTTAGGCGCTTTGCCAGGAGATGCCGCTGCTGGAGTGACTG 995
QY 335 GYGPQQTSG--RGLGGGAG-----AAAAAAGG-----AGQG 370

Db 996 CTGGAGGGCCAGGAGGGAGCCCGTCATCTATCATGGCCATACCCTCACCTCAAGAT 1055
Qy 371 -----GYGLGSGTSGRGLGGGAG----- 392
Db 1056 TCTCTCCGGACGTGGCCCAAGCCGTGCGGACCATCCCTTCAGCTGTCTCCCTTACCC 1115
Qy 393 -----AAAAAAGAGAGGCGGYGGLGSGTSGRGLGGGGA----- 429
Db 1116 TGTCTATCTATCTCTGAGACCACTCGGGCTGGAGCAGCA--GGCTGCCATGGCCGCG 1173
Qy 430 -----GAAAAAAGAGAGGCGGYGGLGSGTSG--RGLGGGAG-----AAAAA 476
Db 1174 CACCTCTGCACCATCTCTGGGGACATGCTGTGTGACACAGGCGCTGGACTCCCCCAATCCC 1233
Qy 477 AAGAGGCG-----YGLGSGTSGRGLGGG--PGQTSRGLGGGAGAAAAA----- 524
Db 1234 GAGGAGTGGCATCCCGAGAGCAGCTGAAGGGCCGGCTCTGTGTGAAGGGAAGAAGCTG 1293
Qy 525 -----AAAAGGAGGCGGYGGLGSGTSGRGLGGGAGAAAAAAGAGAGCGGYGLG 578
Db 1294 CCCGCTCTCGGACGAGGATG--GCCG--GGCTGTGCGGATCGGGAGGGGGAGAG 1350
Qy 579 SGTSGRGLGGGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGGGAGAAAAA 638
Db 1351 --GATGACGAGGAGGAAGAGAGGCTGGAGGCTGCAGCGCAGAGCGGCTGCCCAAGC 1408
Qy 639 AAA-----AGGAGCG-----YGLGSGTSGP 662
Db 1409 AGATCTCCCGGAGCTGTGGCCCTGGCTGTGACTG 1445

RESULT 13

US-09-879-957-37
; Sequence 37, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; KAY, Brian K.
; FOWLKES, Dana M.
; MCCONNELL, Stephen J.
; HOPFMAN, No. US20020034755A1h
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-879-957-37

Query Match 18.7% Score 652; DB 10; Length 1400;
Best Local Similarity 28.3%; Pred. No. 5.2e-31;
Matches 274; Conservative 24; Mismatches 325; Indels 344; Gaps 32;

Qy 14 ASGRGGLGGGCGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGGGCGAGAAAAA 73
Db 2 AAAGGAGGAGAGAGTGTCAAAA--GNAGGATGGCGA-----GGAAAAAGGCACACA 50
Qy 74 AGGAGGCGYGGGLGSGTSGRGLGG-----OGAGAAAAAAGAGAGCAG 118
Db 51 GGAAGCACAGACAGCTGGTTCGCTTTCCATCAACACCAAGAACCCAGCTAAGCCA-- 108
Qy 119 GYGGLGSGTSGRGG-----LGGQ-----GAGAAAAA----- 145
Db 109 -----GCTGTCCAGGCACCCCTGGTCCACTGCAGAAAAAAGGTCCACTTACCACTTCTCCA 163
Qy 146 -----AAAAAGAGCGGYGGLGSGTSGRGG-----YGPQQTSGRGG---L 184
Db 164 CAGAAAAATGTAAAGTGGTGTATTACCGGCACCTGTACCCCTTTGATCCAGAAAGCCAT 223
Qy 185 GGCGAGAAAAAAGAGAGCGGYGGLGSGTSGRGGI-GGCGAGAAAAAAGAGAGCAG 244
Db 224 GATCAATCACTATCCAGCCAGAGACATAGTCATGTTGGATGAAGCAACCACTGGAGAA 283
Qy 245 GYGGLGSGTSGRGLGGGAGAAAAAAGAGAGCGGYGGLGSGT-----SGRG 296
Db 284 CCCG--GCTGGCTTGAGGAGAAATTAAGAGAAAGACAGG---GTGGTTCCCTGCAAACT 338
Qy 297 GLGGGAGAAAAAAGAGAGCGGYGGLGSGTSGRGGYGPQQTSGRGLGGGCGAGAA 356
Db 339 ATGCAGAGAAAAATCCAGAAAAATGAGTTCCCGCTCCAGTGAACCACTGACTGATTCNA 398
Qy 357 AA-----AAAAAGAGAGCGGY--GLGS-----QGT----- 380
Db 399 CATCTGCCCTCCCGCCCAAACTGGCTTGCTGAGACCCCGCCCTTTGGCAGTAACCT 458
Qy 381 -----SGRG-----GLGG----- 388
Db 459 CTTACAGAGCCCTCCACGACCCTTAATAACTGGCGCGACTTTCAGCTCCACGTGGCCACCA 518
Qy 389 -----QGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGG-----QGAGAAAAA 438
Db 519 GCACGAATGAGAAACAGAACGGATAACTGGGATGCTATGGCAGCCAGCCCGCTCTCTCA 578
Qy 439 AAG-----GAGCGGYGGLGSGTSGRGLGG-----AGAAAAA----- 477
Db 579 CCGTTCCAAAGTGGCCGC--CAGTTAAGCGAGAGTCCCGCTTTACTCCAGCCACGCCACT 637
Qy 478 -----AGGAGCGGYGGLGSGTSGRGGYGPQQTSGRGG--- 511
Db 638 GGCTCCTCCCGCTCTCTGTCTAGGCCAG--GGTGAAGAGTGGAGGGCTACAGCTC 695
Qy 512 -----LGGGAGAAAAA----- 528
Db 696 AAGCCCTATATCTTTGGAGAGCCAAAAAAGACAAACCACTTAATTTTAAACAAAATGATG 755
Qy 529 -----GGAGCGGYGGLGSGTSGRGLGGGAGAA---AAAAAAGGAG 571
Db 756 TCATCAGCGTCTCGAAGACAGACATGTGGTTGGAGAGTTCAAGGTGAGAGT 815
Qy 572 GYGGLG-----SGGT-----SGRGLGGGCGAGAAAAAAGAG 608
Db 816 GTTGTTCCTCCCAAGCTCTTACGTGAACACTCATTTTCAGGCCCATTAAGGAAGTCTACAGCA 875

1	PRIOR APPLICATION NUMBER: 60/063541	
2	PRIOR FILING DATE: 1997-10-28	
3	PRIOR APPLICATION NUMBER: 60/063544	
4	PRIOR FILING DATE: 1997-10-28	
5	PRIOR APPLICATION NUMBER: 60/063564	
6	PRIOR FILING DATE: 1997-10-28	
7	PRIOR APPLICATION NUMBER: 60/063734	
8	PRIOR FILING DATE: 1997-10-29	
9	PRIOR APPLICATION NUMBER: 60/063870	
10	PRIOR FILING DATE: 1997-10-31	
11	PRIOR APPLICATION NUMBER: 60/064103	
12	PRIOR FILING DATE: 1997-10-31	
13	PRIOR APPLICATION NUMBER: 60/065311	
14	PRIOR FILING DATE: 1997-11-13	
15	PRIOR APPLICATION NUMBER: 60/066120	
16	PRIOR FILING DATE: 1997-11-21	
17	PRIOR APPLICATION NUMBER: 60/066466	
18	PRIOR FILING DATE: 1997-11-24	
19	PRIOR APPLICATION NUMBER: 60/066772	
20	PRIOR FILING DATE: 1997-11-24	
21	PRIOR APPLICATION NUMBER: 60/069335	
22	PRIOR FILING DATE: 1997-12-11	
23	PRIOR APPLICATION NUMBER: 60/069425	
24	PRIOR FILING DATE: 1997-12-12	
25	PRIOR APPLICATION NUMBER: 60/069870	
26	PRIOR FILING DATE: 1997-12-17	
27	PRIOR APPLICATION NUMBER: 60/068017	
28	PRIOR FILING DATE: 1997-12-18	
29	PRIOR APPLICATION NUMBER: 60/077450	
30	PRIOR FILING DATE: 1998-03-10	
31	PRIOR APPLICATION NUMBER: 60/077632	
32	PRIOR FILING DATE: 1998-03-11	
33	PRIOR APPLICATION NUMBER: 60/077649	
34	PRIOR FILING DATE: 1998-03-11	
35	PRIOR APPLICATION NUMBER: 60/078886	
36	PRIOR FILING DATE: 1998-03-20	
37	PRIOR APPLICATION NUMBER: 60/078939	
38	PRIOR FILING DATE: 1998-03-20	
39	PRIOR APPLICATION NUMBER: 60/079664	
40	PRIOR FILING DATE: 1998-03-27	
41	PRIOR APPLICATION NUMBER: 60/079786	
42	PRIOR FILING DATE: 1998-03-27	
43	PRIOR APPLICATION NUMBER: 60/080107	
44	PRIOR FILING DATE: 1998-03-31	
45	PRIOR APPLICATION NUMBER: 60/080194	
46	PRIOR FILING DATE: 1998-03-31	
47	PRIOR APPLICATION NUMBER: 60/080327	
48	PRIOR FILING DATE: 1998-04-01	
49	PRIOR APPLICATION NUMBER: 60/080333	
50	PRIOR FILING DATE: 1998-04-01	
51	PRIOR APPLICATION NUMBER: 60/081049	
52	PRIOR FILING DATE: 1998-04-08	
53	PRIOR APPLICATION NUMBER: 60/081070	
54	PRIOR FILING DATE: 1998-04-08	
55	PRIOR APPLICATION NUMBER: 60/081195	
56	PRIOR FILING DATE: 1998-04-09	
57	PRIOR APPLICATION NUMBER: 60/081838	
58	PRIOR FILING DATE: 1998-04-15	
59	PRIOR APPLICATION NUMBER: 60/082568	
60	PRIOR FILING DATE: 1998-04-21	
61	PRIOR APPLICATION NUMBER: 60/082569	
62	PRIOR FILING DATE: 1998-04-21	
63	PRIOR APPLICATION NUMBER: 60/082704	
64	PRIOR FILING DATE: 1998-04-22	
65	PRIOR APPLICATION NUMBER: 60/082797	
66	PRIOR FILING DATE: 1998-04-22	
67	PRIOR APPLICATION NUMBER: 60/083322	
68	PRIOR FILING DATE: 1998-04-28	
69	PRIOR APPLICATION NUMBER: 60/083495	
70	PRIOR FILING DATE: 1998-04-29	
71	PRIOR APPLICATION NUMBER: 60/083496	
72	PRIOR FILING DATE: 1998-04-29	
73	PRIOR APPLICATION NUMBER: 60/083499	

1	PRIOR FILING DATE:	1998-04-29
2	PRIOR APPLICATION NUMBER:	60/083559
3	PRIOR FILING DATE:	1998-04-29
4	PRIOR APPLICATION NUMBER:	60/084366
5	PRIOR FILING DATE:	1998-05-05
6	PRIOR APPLICATION NUMBER:	60/084414
7	PRIOR FILING DATE:	1998-05-06
8	PRIOR APPLICATION NUMBER:	60/084639
9	PRIOR FILING DATE:	1998-05-07
10	PRIOR APPLICATION NUMBER:	60/084640
11	PRIOR FILING DATE:	1998-05-07
12	PRIOR APPLICATION NUMBER:	60/084643
13	PRIOR FILING DATE:	1998-05-07
14	PRIOR APPLICATION NUMBER:	60/085573
15	PRIOR FILING DATE:	1998-05-15
16	PRIOR APPLICATION NUMBER:	60/085579
17	PRIOR FILING DATE:	1998-05-15
18	PRIOR APPLICATION NUMBER:	60/085580
19	PRIOR FILING DATE:	1998-05-15
20	PRIOR APPLICATION NUMBER:	60/085582
21	PRIOR FILING DATE:	1998-05-15
22	PRIOR APPLICATION NUMBER:	60/085700
23	PRIOR FILING DATE:	1998-05-15
24	PRIOR APPLICATION NUMBER:	60/086023
25	PRIOR FILING DATE:	1998-05-18
26	PRIOR APPLICATION NUMBER:	60/086392
27	PRIOR FILING DATE:	1998-05-22
28	PRIOR APPLICATION NUMBER:	60/086486
29	PRIOR FILING DATE:	1998-05-22
30	PRIOR APPLICATION NUMBER:	60/087098
31	PRIOR FILING DATE:	1998-05-28
32	PRIOR APPLICATION NUMBER:	60/087208
33	PRIOR FILING DATE:	1998-05-28
34	PRIOR APPLICATION NUMBER:	60/087609
35	PRIOR FILING DATE:	1998-06-02
36	PRIOR APPLICATION NUMBER:	60/087759
37	PRIOR FILING DATE:	1998-06-02
38	PRIOR APPLICATION NUMBER:	60/087827
39	PRIOR FILING DATE:	1998-06-03
40	PRIOR APPLICATION NUMBER:	60/088025
41	PRIOR FILING DATE:	1998-06-04
42	PRIOR APPLICATION NUMBER:	60/088028
43	PRIOR FILING DATE:	1998-06-04
44	PRIOR APPLICATION NUMBER:	60/088029
45	PRIOR FILING DATE:	1998-06-04
46	PRIOR APPLICATION NUMBER:	60/088033
47	PRIOR FILING DATE:	1998-06-04
48	PRIOR APPLICATION NUMBER:	60/088167
49	PRIOR FILING DATE:	1998-06-05
50	PRIOR APPLICATION NUMBER:	60/088202
51	PRIOR FILING DATE:	1998-06-05
52	PRIOR APPLICATION NUMBER:	60/088212
53	PRIOR FILING DATE:	1998-06-05
54	PRIOR APPLICATION NUMBER:	60/088217
55	PRIOR FILING DATE:	1998-06-05
56	PRIOR APPLICATION NUMBER:	60/088326
57	PRIOR FILING DATE:	1998-06-04
58	PRIOR APPLICATION NUMBER:	60/088655
59	PRIOR FILING DATE:	1998-06-09
60	PRIOR APPLICATION NUMBER:	60/088722
61	PRIOR FILING DATE:	1998-06-10
62	PRIOR APPLICATION NUMBER:	60/088738
63	PRIOR FILING DATE:	1998-06-10
64	PRIOR APPLICATION NUMBER:	60/088740
65	PRIOR FILING DATE:	1998-06-10
66	PRIOR APPLICATION NUMBER:	60/088811
67	PRIOR FILING DATE:	1998-06-10
68	PRIOR APPLICATION NUMBER:	60/088824
69	PRIOR FILING DATE:	1998-06-10
70	PRIOR APPLICATION NUMBER:	60/088825
71	PRIOR FILING DATE:	1998-06-10
72	PRIOR APPLICATION NUMBER:	60/088826
73	PRIOR FILING DATE:	1998-06-10

```

; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908

Query Match          17.4%; Score 609; DB 12; Length 4440;
Best Local Similarity 26.0%; Pred. No. 3.3e-28;
Matches 283; Conservative 27; Mismatches 332; Indels 446; Gaps 41;

QY 15 SGRGGLGG----QGAGAAAAAAGAGAGQG---GYGGLGSGQT-SGRGGLGGGAGAA 66
Db 152 TGTGTCGCAATTGTGTATGTCATGTCATGTGTGCGTGACACGCTGTGTGCGTGA 211

QY 67 AAAAAAGAGAGGQGGYGGGLGSGQTSRGLGGGAGAAAAAAGAGAGGQGGYGLGSGQ 126
Db 212 CATGTGTCATGTGT---GTGTTGT-GTGCTGTGTGTGCACATGTAGAAA-GAAAGTGAT 265

QY 127 GTSRGLGGGQAGAAAAAAGAGAGAGGQGGYGLGSGQ--TSGPGYGPQQTSRGLG 184
Db 266 GTGTGGCAATGGAGAAATTCACCCAGAGAGACTGTGCTGGGCTGCCACGAGAAAT 325

QY 185 GGQGA-----GAAAAA-----AAAAAGAGGQGGY-----GLGSGQT 216
Db 326 GGTGACACTTCGAGAAACGGGTTCCAAATGCAGTTGCGCTTCCAGTCTCTGTGTGACCT 385

QY 217 SGRGGLGG-----QGAG-----QGAG-----AAAAA 233
Db 386 TGGGCAAGTCACTTCAGCTCCCTGAGCTTTGTTTTAAAAATATTTTTTAAATGTATAA 445

QY 234 AAAAAAGAGGQGGYGLGSGQTSRGLG---GQAGAAAAAAGAGAGAGAGAGAGAGAG 281
Db 446 ACCATGGACCATTACATATGAGAGAGAAATGTGTGTGCAACATTCAGTTAATAACACAA 505

QY 282 GGYGGLGSGQT-----SGRGLGG---QGAGAAAAAAGAGAGAGAGAGAGAGAGAG 319
Db 506 GGTGGAGGAGTGCTGCTCAGACCCAGAGCTGACACAGGGAACCTCCCCAGAGGCTGCA 565

QY 320 GG-----YGGLGSGQTSRGLG-----YGGLGSGQTSRGLG-----PGQTSRGR 346
Db 566 GGGGCTTCCCTCCCAACCTCCAAAGGCGCCACTGTGTGAGCGCGCTGTATCTGTTCAGTG 625

QY 347 GLGGQG-----AGAAAAA-----AAAAAGAGQ-GYG-----GLG 376
Db 626 GAATGCCATGTGCCACGTTCTCAGNAGTAAGCCACTAGGTAGTGTGACACTGTGTG 685

QY 377 -----SQTSRGLG-----QGAG-----QGAG-----AAAAAAGAG 405
Db 686 CCCAGCTTGTGCAAGCCCCCTTGTGTGTCTATCTTATTGACCTCCCAATAGCATGCT 745

QY 406 GGGG-YGGLGSGQT-----SGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 442
Db 746 AAGGTCACTGTCTCACTTCCCTCCCTTTTGAAGATGAGGAGACACAAATCTTAGATGA 805

QY 443 AGQGGYGLGSGQTSRGLG-----LGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 478
Db 806 AATGGAGTCCAGTGGAAATAGGATCCAGACAGATTAATCCATCTCAAGCCTGAAT 865
```

```

QY 479 -----GGAG----- 482
Db 866 TCTTCCATCCAGCTACGCTTGAAGCTCAATCTCTTCTCTCTCTCTCTCTCTCTCTCT 925
QY 483 -----QGYGG--LGSQGTSPGG-YGPGQQTSGR 509
Db 926 CCCACCCCCAGATATATCCCATCGCTGTGTGTGGACAGTAGCCATGACTGGGTTTGG 985
QY 510 GGLGGGAGAAAAAAGGAGGQGGYGLGSGQT-----SGRG 548
Db 986 TAAAGTTGCTGAATATCAGGCTGTGTGTAGTTTTCATTTTCCACTTCCAGTGAA 1045
QY 549 GLGGGAGAAAAAAGGAGGQGGYGLGSGQTSR-----GGLGGGAGAAAA 597
Db 1046 ATGGGCCCCCATGAAAAAGSCA-GCTCAAGTTCTAATTTACTCAAAGGAAGACAAAG 1104
QY 598 -----AAAAAAGGA---GQGGYGG----- 614
Db 1105 GTCTTCTGTTCACCTACCCCTAAGGATTGGGCTAGACACTGGGAATTTACTAATTATG 1164
QY 615 -----LGSQGTSGRGLGGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 655
Db 1165 AATTCAGTGTCTTCTCTTCTGTAAGAGAGGCGTGGAATCAACGCTGAGTGAAGGCATCA 1224
QY 656 QGTSRPGG 663
Db 1225 AGTTTAAAG 1232
```

Search completed: December 18, 2002, 16:39:01
Job time : 20.4075 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 20,1216 Seconds
(without alignments)
3253.588 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHGSMASRGGL.....GGYGPQQTSGITRPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2135	61.1	718	2 A36068	major ampullate fi
2	1386.5	39.7	2639	2 T31328	fibroin - Chinese
3	1277.5	36.6	1901	2 F70806	hypothetical glyci
4	1254.5	35.9	1489	2 D70807	hypothetical glyci
5	1228.5	35.2	1079	2 B70807	hypothetical glyci
6	1197	34.3	1306	2 A70934	hypothetical glyci
7	1182.5	33.9	1381	2 E70806	hypothetical glyci
8	1173	33.6	1660	2 A70869	hypothetical glyci
9	1157	33.1	1538	2 H70846	hypothetical glyci
10	1145	32.8	1329	2 E70917	hypothetical glyci
11	1139	32.6	853	2 A70896	hypothetical glyci
12	1135	32.5	627	2 A44112	hypothetical glyci
13	1114.5	31.9	749	2 A70812	spidroin 2, dragli
14	1106.5	31.7	914	2 H70987	hypothetical glyci
15	1075	30.8	882	2 F70812	hypothetical glyci
16	1044.5	29.9	778	2 F70963	hypothetical glyci
17	1036	29.7	767	2 E70895	hypothetical glyci
18	1028	29.4	957	2 D70835	hypothetical glyci
19	1024	29.3	837	2 E70835	hypothetical glyci
20	1014	29.0	783	2 F70824	hypothetical glyci
21	1004	28.7	731	2 C70974	hypothetical glyci
22	988	28.3	801	2 F70824	hypothetical glyci
23	984.5	28.2	714	2 A70807	hypothetical glyci
24	984	28.2	741	2 G70917	hypothetical glyci
25	976	27.9	860	1 EAMS	hypothetical glyci
26	968.5	27.7	1011	2 F70620	hypothetical glyci
27	939	26.9	864	1 EART	elastin precursor
28	923.5	26.4	923	2 E70820	hypothetical glyci
29	922.5	26.4	667	2 A70893	hypothetical glyci

30	903.5	25.9	694	2 F70868	hypothetical glyci
31	893	25.6	484	2 G70846	hypothetical glyci
32	869	24.9	603	2 A70770	hypothetical glyci
33	860	24.6	1585	2 T31611	hypothetical prote
34	853.5	24.4	465	1 S01820	glycine-rich cell
35	833.5	23.9	576	2 A70900	hypothetical glyci
36	817.5	23.4	584	2 G70804	hypothetical glyci
37	808.5	23.1	618	2 A70989	hypothetical glyci
38	804.5	23.0	770	2 S59623	hypothetical glyci
39	804	23.0	639	2 D70931	tropoelastin - she
40	800	22.9	591	2 B70523	hypothetical glyci
41	799	22.9	606	2 H70816	hypothetical glyci
42	784.5	22.5	747	1 EABO	elastin precursor,
43	776.5	22.2	615	2 F70589	hypothetical glyci
44	774.5	22.2	588	2 H70971	hypothetical glyci
45	763	21.8	641	1 Q0BE31	nuclear antigen EB

ALIGNMENTS

RESULT 1

A36068

major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)

C:Species: Nephila clavipes

C>Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997

C:Accession: A36068

R:Xu, M.; Lewis, R.V.

Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990

A:Title: Structure of a protein superfiber: spider dragline silk.

A:Reference number: A36068; MUID:90384959; PMID:2402494

A:Accession: A36068

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-718 <XUA>

A:Cross-references: GB:M37137; NID:gl59711; PID:gl59712

A>Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue

Query Match 61.1%; Score 2135; DB 2; Length 718;

Best Local Similarity 68.3%; Pred. No. 1.1e-112;

Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY	23	QAGAGAAAAAAGAGAGGCGGGLGSGQT--SGRGLGGGAG--AAAAAAGAGNGQ	79
DB	1	QAGAG-----AAAAAGAGGCGGGLGGGAGGCGGCGGAGAGAGAGAGAGAGGAGG	56
QY	80	GGYGGGLGSGQTSRGLGGGCGAGAAAAAAGAGAGGCGGCGGGLGSGQTSRGLGGGCGA	139
DB	57	GGYGGGLGSGQ-----AGRGGAG-----AAAAAGAGGCGGCGGGLGSGQ-AGRGLGGGGA	107
QY	140	GAAGAAAAAGAGAGGCGGGLGSGQTSRGLGGGCGAGAGAGAGAGAGAGAGAGAGAGAG	199
DB	108	G-----AAAAAGAGGCGGGLGNGQ-----AGRGGQ-----AAAAA	142
QY	200	AGGAGGCGGGLGSGQTSRGLGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	250
DB	143	AGGAGGCGGGLGSGQ-AGRGLGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	201
QY	251	SQTSRGRGLGGGCGAG	289
DB	202	SQL-AGRGLGGGCGAG	260
QY	290	QGTSGRGLGGGCGAG	348
DB	261	Q-----AGRGEGAG-----AAAAAGAGGCGGCGGGLGSGQ-AGGCGGGLGSGQAGRGL	311
QY	349	GGGAG	390
DB	312	GGGAG	370
QY	391	AGAAAAAAGAGAGGCGGGLGSGQTSRGLGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	450
DB	371	AG---AVYAAAAGAGAGGCGGGLGSGQ-----AGRGGAG-----AAAAAGAGAGAGAGAG	419

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: B70807
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1079 <COL>
A: Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17749.1; PID:g292444
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv3512
C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
Query Match 35.2%; Score 1228.5; DB 2; Length 1079;
Best Local Similarity 40.0%; Pred. No. 5e-62;
Matches 318; Conservative 25; Mismatches 303; Indels 149; Gaps 29;
QY 14 ASGRGGLGGCGACAAAAAAGGACGGYGGGLGSGQTSRGLGGGAGAAAAA 73
DB 38 AGGAGAGGTG-CTGGAAGTGTGGGQNGGNGG--GTGGKGTGTDGALAGSSGGAGG 94
QY 74 AGG-----AGGGYGGGLG-----SOCTSGRGLGGQ 99
DB 95 KCGNGGDACKAGTGSAPGTAGTGGDGKCGNGGIGAGTTGPTGASGTTGSGGAGGT 154
QY 100 GA-GAAAAAAGGA-----GOGGYGGLG-----SOCTSGRGLGGQ----- 137
DB 155 GGDGAANGGTAGAGAGNGGKGGGAGTSTAGNSGAGGSGGKGDAGAGGAGAT 214
QY 138 -----GAGAAAAAAGGAGGAGGAGGGLGSG-----GTSGPGGYGPGQOTSG 180
DB 215 PGANGTAGNGDGDGAAGAVTSGATGADGGHGTGAAGNGGTGGAGGIDYGGG 274
QY 181 RGLGLGGGAGAAAAAAGAGAG-OGGYGGLGSG-----GTSGRGLGGQ 225
DB 275 TGTGTGNGNGALCGGAGDAGSNGSGNGIGCKGNAGAGGANGTGVANGTGGD 334
QY 226 G--AGAAAAAAGAGAGGGYGGGLGSGQTSRGLGGQ-----AGAAAAA 272
DB 335 GNGGAAGAATAGSNGAGTGSAGNG--GTGGRGSGGAGGDIQVGGKGGNGADGE 392
QY 273 AAAGGAG-----OGGYGGLGSGQTSRGLGGGAGAAAAAAGGAGGAGGAGG 325
DB 393 VGGAGGAGGSPNTSPGNGGGGQGGSGGAG--GAAGAGGAGGANGTAGNGGGGAGGT 451
QY 326 GSQCTSGP---GGYGPQOOTSRRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGG 382
DB 452 GGAGAAASSATNGSGGAGGTGGDGGSGGAGTGGAGTGAAGDGGGQGGAGG--GAGG 510
QY 383 RGLGGQGG-----AGAAAAAAGG--AGGGYGGGLGSG--GTSGRGLGGQGA 429
DB 511 QCGAGGAGGTGGNGNITGTTAGTAGAAGNGAAGKGGAGGQGTGGTGGGQGGAGDGG 570
QY 430 GAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 489
DB 571 AGTGGDRTVGGGTVPAGSGGQ--GNAGGGGAGG--GGADGSGGDDAGTGGNGN 626
QY 490 G-----SQGTSGPGGYGPGQOTSRRGLGGGAGAAAAAAGGAGGAGGAGGAGG 542
DB 627 GGNRNSGNTGAGAGNGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686
QY 543 GTSRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 597
DB 687 GNGNGGAGTAGMGNAGGAGTGSNGNGGSGGNGGNGAGMGNAGTSGSDGGAGNGGAAGT 746
QY 598 AAAAAAGGAGCGGCGGLGSGQTSRRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGG 657

DB 747 GGTGGDGLTGTGTTGGSG--GTGGDGGNGGADNTANMTAQAGDGGNGGCGFGGA 804
QY 658 TSGPGGYGPGQOTSG 672
DB 805 GAGGGGLTAGANGTG 819
RESULT 6
A70934
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70934
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: A70934
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1306 <COL>
A: Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAAL17449.1; PID:g290
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv0578c
C: Superfamily: collagen alpha 1(IV) chain
Query Match 34.3%; Score 1197; DB 2; Length 1306;
Best Local Similarity 38.1%; Pred. No. 3.2e-60;
Matches 325; Conservative 38; Mismatches 296; Indels 194; Gaps 35;
QY 3 GSHHHHHG-----SWASRRGLGGGAGAAAAAAGG-----AGGGYGGGLGSGQTS 53
DB 388 GSAHANGANGAPTSTSGNGDGGKADAISSGTGANGRGDGGGQVNGGAG--GAG 445
QY 54 GRGGLGGGAGAAAAAAGG--AGOGYGG--LGSQTSRGLGGQGA----- 101
DB 446 GRGAGGLGFGSAPRPGAGGTGGAGNGGTQAGDGTGAGGAGDGGSGGAGSIGF 505
QY 102 GAAAAAAGG-----AGOGYGG-----GSQTSRGLGGQGA----- 139
DB 506 NASAPGAAGSPGNGNGGPGGAGGAGGAGGALAAASQNGSQGAGSDGAGNGGTGPN 565
QY 140 ---GAAA-----AAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 186
DB 566 GCHGAAGALGVNGVGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 625
QY 187 QGAGAAAAAAGG-----AGOGYGGGLGSG-----GTSGRGG 221
DB 626 RGADATGCTGAGSGRGGDGLVNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 685
QY 222 LGGGAGAAAAAAGG--AGOGYGGGLGSG-----GTSGRGGGLGGQ-- 264
DB 686 AGSGGAGAGNGGTGGAGGTGGYGGTGGGSDGVNGSSAGADGHPGTGGVGTGGKGGD 745
QY 265 AAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 323
DB 746 GCDGGAAPNGVAGSQGPGGAGCDG--GTGGVGGNGRGIDCADCATAGAGDGGAGGAG 803
QY 324 GLGSG--GTSPPGGYGP-----GOOTSRRGLGGQ----- 352
DB 804 GKGRGTGPGGAGPAGTTGSGAGNGSGSGGTGDPDGGNGANGSVFTNNGIGNGG 863
QY 353 -AGAAAAAAGGA-----GOGGYGGGLGSGQTSRRGLGGGCGAGAA 394
DB 864 NGNAGPSGAGGSGGAGTGTGATGSSSSIHVNGNGNGNGNDHALSGNCAAGNGNGG 923
QY 395 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450

QY 124 -----GSGTSGRGLGQGA-----GAAAAAA 147
Db 630 AFKFDGGAGGDPGAGCGKAGAGATEGVTGATGATVHSGNGKGGNCADATVAG 689
QY 148 AAAGAGGQGGGL-GSQGTSGPGYG-----PG---QQTSGRG 182
Db 690 ANGKGGAGNGGLVDGGAGGDDGGGGAAGANGANVGDGDTLSGQPGGSEANGGQ 749
QY 183 GLGGGAGAAAAAAGAGAGGQGGYGGGLGSGQTSRGLGGGCGAAAAAAGAGAG 242
Db 750 GVGGGAGGAGGAGGAGSALGSGNGRGDAGQAG----GAGGAGGAGGAGSVSDGG 805
QY 243 QGGYGLGSGQTSRGLGGGCGAAAAAAGGA-GGGYGLGSGQTSRGLGG 301
Db 806 PGKGGAGGAGGAGS-GGGKAGSADSAEAVGAGGKGGDGVG--GVGGDGGPGD 862
QY 302 G-AGAAAAAAGAGAGGQGGYGGI-GSQGTSGPGYG-----PGQQTSGRG 347
Db 863 GGAGNAPAGVGHGVGGVGGGLGGAGNGDGGHGGDGGDGGDGGDGGDGG 920
QY 348 LGG-CCAGAAAAAAGAGAGAGGQGGYGGGLGSGQ-----GTSR---GG 385
Db 921 LGGDSNGTRAASGVADSDHDPGSGNGGNGGNGAGAGAGGAGGAGGAGGAGGAG 980
QY 386 LGGGAGAAAAAAGAGAGAGGQ-----GGYGLGSGQ-TSGRGLGGQ-GAGAAAA 435
Db 981 AGCNGGDAAGANGANGSCAPGSDALALQPGNGGCGDAGAGGAGGAGGAGGAG 1040
QY 436 AAAAAAGGAGGQGGYGLGSGQTSR-----GGLGGGCGAAAAAAGAGAGGAG 486
Db 1041 DGGAGNGAGNGGAGGAGGAGGAGAGANGIDSIGTGGAGGAGGAGGAGGAGGAG 1100
QY 487 GG-----LGSQGTSGPGYG-----PGQQTSG 507
Db 1101 GCAAPSGTVGSHGTGSGVGGDGLGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1158
QY 508 GRGGLGGQ-----GAGAAAAAAGAGAGGAGGQGGYGLGSGQ 543
Db 1159 GRGGLGGSGNGTSAANGYDASKHPLTGGDGGVGGNGAKAAAAAGDGGQGGD--GNAG 1216
QY 544 TSGRGLGGGAGAAAAAAGGA-GGGYGG-----LGSQGTSGRGLGGQGA----- 593
Db 1217 LFGDGGAGDGTAAEALGDDGAGGAGGAGGAGDAGDAGDGGGKGGDGAHGLG 1276
QY 594 -----GAAAAAAGGA--GGGGYGLGSGQTSRGLGGGCGAGAA-AAAAAAGG 644
Db 1277 VAGNGGAGGAGGAGGAGGAGFLGDSNGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336
QY 645 AGGGYGLGSGQTSRGGYGPQQTSG 672
Db 1337 NGGGGTGGGNGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1364

RESULT 9
H70846

hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1538 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g326151
A:Experimental source: strain H37RV
C:Genetics:

A:Gene: Rv3345c

C:Superfamily: collagen alpha 1(IV) chain

Query Match 33.1%; Score 1157; DB 2; Length 1538;

Best Local Similarity 34.4%; Pred. No. 6.2e-58;

Matches 317; Conservative 36; Mismatches 308; Indels 260; Gaps 31;

QY 1 MRGSHHHHHHSGMASRGGGLGGGCGAGAAAAAAGG-----AGQGGYGLGSGQTS 53
Db 549 LNSGACANGTAPTISGNGGNGGAGATPTVAGENGAGGNGHGGSVGNGAGGAGNGVA 608
QY 54 G-----RGLGGGGA-GAAAAAAGAGAGGQGGYGLGSGQ-----GTS 91
Db 609 GTGLALNGGNGGNGGIGGNGGSAAGTGGDGGKGGNGAGANGDFASANGANGGQGGNG 668
QY 92 GRGGLGGGCGAAAAAAGAGAGGQGGYGLGSGQTSRGLGGGGA----- 139
Db 669 GNGGIGGKGDAFATPTAKAGNGGAGNG-----GNVAVAGGGAGGKGAIPAMKATGADG 724
QY 140 -----GAAAAAAGAGAGGQGGYGG-LGSQGTSGPGYGPQO-----T 178
Db 725 TAPTSGDGGNGNGASPTVAGNGDGGKGGSGGNGVNGGAGGAGNGAAGQAGTPGPT 784
QY 179 SRGGLGGGCGAAAAAAGG-----AGQGGYGLGSGQTSRGLGGGCGAAAAAAG 236
Db 785 SDDSSTSGTDGAGGNGGAGAGGTLAGHGGNGKG--GNGGQGGTGGAGRGADGAGPN 842
QY 237 AAGGAGGQGGYGLGSGQTSRGLGGQ-CAGAAAAAAGAGAGGAGGAGGAGGAGGAG 286
Db 843 ANGANGENG-----GSGGNGDGGAGGAGGAGKQAAGYTDGATGTGSDGNGDGGKAG 898
QY 287 -----LGSQGTSGRGLGGQ----- 301
Db 899 DGGAGENGLNSAMLPGGCTVGNPCTGNGGNGAGVCGTGGKAGTGLTGLDGTGIT 958
QY 302 -----GAGAAAAAAGG-----AGQGGYGG-----LGSQGTSG 331
Db 959 PNGGNGGNGGNGKGTAGNGSAGAGGNGGSGLNGDAGNGGNGGALNAQAGFFGTG 1018
QY 332 GPGGY---GPQQTSRGLGGGCGAAAAAAGAGAGGQGGYGLGSGQTSRGLGG 387
Db 1019 GKGNGGNGGAGWINGLGGFGGAGGGAVDVAATTGGAGGNGGAGGFASTGLGPGGAG 1078
QY 388 GOGAGAAAAAAGGA---GQGGYGLGSGQ-----TSGRGLGGGCGAGAAAAAAG 436
Db 1079 GPGAGDFASGVGGAGGDDGGAGSVGGFCGGGTGGERTGNGSGDGGGGLSLOG 1138
QY 437 AAAAG-----AGQGGYGLGSGQTSRGLGGGCGAGAAAAAAGAGAGGAG 484
Db 1139 NGGLGGNGGVSETGFGAGNGGYPG--GPEGNGLGGNGGAGGNGGVSTTGGDGGAG 1196
QY 485 GYGLGSGQTSRGGYGPQQTSRGLGGGCGAGAAAAAAGG-----AG 532
Db 1197 GKGNGGDD--GNVLGGDAGSGGAGNGGIGTDAGGAGGAGGAGGNGSSSTTTGNAG 1254
QY 533 QGGYGLGSGQTSRGLGGGCGAGAAAAAAG-----AAAGAGGQGGYGLGSGQTS-GRGGL 588
Db 1255 SGAGGNGGTGLNAGGAGGAGGAGVAGVAFNAVGGDGGNGGNGHGGDGTGGAGCK 1314
QY 589 GQGGAGAAAAA-----AAAAGAGGQGGYGLGSGQTSRGL-----GGGAGAAA- 635
Db 1315 GNGSSGAAAGSVVNVYTAGHGGNGGNGGNGSAGAGGQGGGAGGAGGNGHGGGATGG 1374
QY 636 -----AAAAAAG-----AGQGGYGLGSGQTS-----G 660
Db 1375 DGGNGGNGGNSNSTGVAGLAGAGAGGNGGTSAAAGHGGGSGSGGTTTGGAGAAAG 1434
QY 661 PGYGP-----QQTSGIRR 675
Db 1435 NGGAGAGGGSLSGTGSGGPRR 1455
RESULT 10

E70917

hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70917
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70917
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1329 <COL>
A:Cross-references: GB:J95844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g2131046
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1450c
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 32.8%; Score 1145; DB 2; Length 1329;
Best Local Similarity 35.9%; Pred. No. 2.6e-57;
Matches 323; Conservative 35; Mismatches 300; Indels 242; Gaps 35;

```
Qy 11 GSMASGRGLGGGCA-----GAAAAAAAAGGA-----GQGGY 45
Db 139 GGIILWNGGAGGAPQGVGGAGAAAGLFTGTGAGGAGGAGAGAGGGGGLNGSVG 198
Qy 46 GLGSOQTSGRGLGGGAGAAAAAAGAGGAGGGYGLGSOQTSGRGLGGGAGAAA 105
Db 199 GAGQSLIG-GATGAGAGNAGLFGVGTGPGGPGGPGVG--GTGAGGLGLYAGG 255
Qy 106 AAAAAAGG-AGQGGYGLGSLQ-----GTSRGLGGGAGAAAAA-----AAAA 150
Db 256 HGGAGGPGPIGVGGHGGVGAAGLLVGGHGGAGGAGGAGVAGAACGDLSPHGTSGVG 315
Qy 151 GGACGGYGLGGS--QGTSPGGYGPQOQTSGRGLG-----GGAGAAAAA 200
Db 316 GDAGDGTGGRGLAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 375
Qy 201 GGAGQ--GGYGLGSOQTS-GRGGLGGGAGAAAAAAGGA-----GQGGYGLG 250
Db 376 GGAGSTTGAHGAAGASPTSGNGGAGGAGGAGHFSGGKAGNGGAGGAGGLVNGGAGAG 435
Qy 251 SQGT-----SRRGLGGGAGAAAAAAGGA--GQGGYGLGSOQTSGR 295
Db 436 GNGAPAPPSSGDPNGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 495
Qy 296 GGL-----GGGAGAAAAA-----AGGA-----GQGGY 322
Db 496 NGLGAGAGDTGDKGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555
Qy 323 GGLGSOQT-----SGPGYGPQOQTSGRGLGGGGA-----GAGG 353
Db 556 GNGAKGTFTNGGDCVGGNGGNGSGRIGGAGGIGGAGGTAGAGCARGATPTSGNGGTG 615
Qy 354 --GAAAAAAGAGAGGGYGL-GSQQTSGRGLGGGAGAAAAAAGGAGQ--- 407
Db 616 GNGANATVAGCAGCAGGKGGNGLVNGGAGGKGGDGAGVAGSPPTAGESGTSQNGG 675
Qy 408 -GGYGLGSLQ-----GTSRGLGGGAGAAAAAAGG-----AGQGGY 448
Db 676 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 735
Qy 449 GGL-----GSQTSRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGG 491
Db 736 GGLAGNLFQNGIUGVGGSGKGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 795
Qy 492 QGTSPGGYGPQO-----TSRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGG 542
Db 796 QGSGGAGSTPGAKAHGFTPTSGDGGDGGNGN-----SQVYGGNGDGGNGGSGSA 850
```

```
Qy 543 GTSRGRGLGGGAGAAAAA-----AAAGGAGGGYGLGSOQTSRGG 587
Db 851 GTGNGRGGGDFAGFGMSANATPNPENGPNPGNGGAGGAGGAGLNG-GNGGAGNGG 909
Qy 588 LGGCGAGAAAAAAGAGGAGQ---GGYGLG-----SQTSGRGLGGQ 629
Db 910 LGFGGNGCAAGANGVAVCAPQPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 969
Qy 630 G-----AGAAAAAAGAGGAGGAGG-LGSQGTSGPGYGPQOQTSGIRPAAKL 680
Db 970 GGAPEGDGANGGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1026
```

RESULT 11

A70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAAL1707.1; PID:el218
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match 32.6%; Score 1139; DB 2; Length 853;
Best Local Similarity 40.3%; Pred. No. 4.1e-57;
Matches 298; Conservative 27; Mismatches 304; Indels 110; Gaps 28;

```
Qy 11 GSMASGRGLGGGAGAAAA---AAAAAGCAGCGYGLGSOQTS-----GRGSL 58
Db 137 GLLYNGNGNGTSTTAGVAGNGGAAGLNGGAGGGGAGAGCGNGAGGAGGAGGAGG 196
Qy 59 GCGAGAAAAAAGGA-----GQGGYGLGSOQTS---RGGLGGGAGAAAAA 107
Db 197 GAGGTSVIPGAVGNGGAGGAGGAGLWGTGGAGGAGGAGGAGGAGGAGGAGGAGG 256
Qy 108 AAAAGAGCGGYGLG-----SQTSGRGLGGGAGAAAAAAGGAGGAGGAGG 159
Db 257 LFGDAGAGNGGKGGAGGAFAFINFTAGDGGAGGAGGAGGAGGAGGAGGAGGAGG 316
Qy 160 GLG--SQTSPGGYGPQOOTS---GRGLGGGGA---GAAAAAAGGAGGAGGAGG 209
Db 317 GAGGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
Qy 210 GLGSOQTSGRGLGGGQ-----AGAAAAAAGGAGGAGGAGGAGGAGGAGGAGG 258
Db 377 GAGDQ--GAGCAGGNARLWGVGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 434
Qy 259 GLGCGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 318
Db 435 GTG--GAGGEGGAAGLLVGTGGHGGDGGAG--GAAVKGGDGGAGGAGGAGGAGGAGG 490
Qy 319 QGGYGLGSOQTSGPGYGPQOQTSGRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGG 371
Db 491 SGGSGDGGCGGAAGPAGW-----LFGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 545
Qy 372 YGSLGSOQTS-----GRGLGGGAGAAAAA-----AAGCAGCGYGLGSOQTSRGG 422
Db 546 NGNGGNGCATGWLIVNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 603
```

Qy	423	GLGQGA--CAAAAAAAAAAGAG--QGYVGLG-----SQGTSRGGLGGQ--GAGAAAAA	474
Db	604	GAGNAGLLGVGAGGAGGASGGAGDRGAGGTCFTSSDGGAGGCGDGGNGGAGGTGELL	663
Qy	475	AAAGAGAGCGYGLGSQGTSGPGGYGPGQQTSGRGLGGGAGAAAAAAGAGAGAGG	534
Db	564	FGAGGCGCGSGGAAIDG--GNGGAGNGGTTDGMGNGGSGGG---AGSGGGGGAGGN	718
Qy	535	YGGLGSGQTSRGRLGGGAGAAAAAAGAGAGCGYGLGSQGTSGRGLGGGAG	594
Db	719	GAWLFGNGGAGGGGKGGNGAG-----GGLGGSGFLPGLNGSGDGDGNGAP	768
Qy	595	AAAAAAAAAGAGAGCGYGLGSQGTSGRGLGGGAGAAAAAAGAGAGCGYGLG	654
Db	769	GGVLYGNNGAG--GGSSSGIGPGAT--GGAGRGKGGDQAQLTGDGNGGNGAGGTG	824
Qy	655	SQGTSGPGYGPQQTSGI	673
Db	825	--GTPGPGCGSGGGLGL	841

RESULT 12

A44112

spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)

N:Alternate names: silk fibroin, dragline

C:Species: Nephila clavipes

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998

C:Accession: A44112; S27824

R:Hinnman, M.B.; Lewis, R.V.

J. Biol. Chem. 267, 19320-19324, 1992

A:Title: Isolation of a clone encoding a second dragline silk fibroin. Neph.

A:Reference number: A44112; MUID:92406876; PMID:1527052

A:Accession: A44112

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-627 <HIN>

A:Cross-references: GB:M92913; NID:g159713; PID:g159714

A:Note: sequence extracted from NCBI backbone (NCBIP:113893)

R:Hinnman, M.B.; Lewis, R.V.

submitted to: The EMBL Data Library, May 1992

A:Description: Isolation of a clone encoding a second dragline silk fibroin

A:Reference number: S27824

A:Accession: S27824

A:Molecule type: mRNA

A:Residues: 19-627 <H12>

A:Cross-references: EMBL:M92913

[illegible]

Db 273 G-----SAAAAAAAAAGPGGYG-----PQQQGGYGPQQ-----GPSCAGS 310

Qy 356 AAAAAAAGAGGAGGCGYGLGSQGTSGRG-----GLGGGAGAGAAAAA-----GGAGGCGY 410

Db 311 AAAAAAGPGQGLGCGYGP-CQQGPGYGPQQQPGYGPASASAAAAAAGPGQGGPGY 369

Qy 411 GGLGSQGTSGRGGLGGCGAGAAAAAAGGAGCGGCGGLGSQGTSG--RGGLGGCGAG 468

Db 370 GP-GQQGPGSGP-----SASAAAAAAGPGGYGP-GQQGPGYAPGQQGPGSGP 417

Qy 469 AAAAAAAGAGGAGGCGYGLGSQGTSPGPGYGPQQGTSGRGGLGGCGAGAAAAA----- 528

Db 418 SASAAAAA--AGPGYG-----PQQQGGYAPQQ-----GPSGPGSAAAAA----- 463

Qy 529 GGAGGGYGLGSQGTSGRGGLGGCGAGAAAAAAGGAGCGGCGGLGSQGTSGRGGL 588

Db 464 --AGPGYGP-AQQGPGP-----TASASAGPGYG-----PAQQGP 500

Qy 589 GGQAGAAAAAAGAGGAGGCGYGLGSQGTSGRGGLGGCGAGAAAAAAGGAGCGG 648

Db 501 AGYGPASAVAASA---GAGSAGYGP-CSQASAAASRLASPDSGARVASVNLVSSGPT 555

Qy 649 GYGGLGSQ-----GTSGGP 662

Db 556 SSAALSSVISNAVSIQASNPG 577

RESULT 13

A07012

Hypothetical glycine-rich protein Rv0833 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: A70812

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Annor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70812

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-749 <COL>

A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17639.1; PID:g291

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: Rv0833

C:Superfamily: elastin

Query Match 31.9%; Score 1114.5; DB 2; Length 749;

Best Local Similarity 39.6%; Pred. No. 8.7e-56;

Matches 303; Conservative 18; Mismatches 309; Indels 135; Gaps 32;

Qy 16 GRGGLGGCGAGAAAAAAGGAGCGGCGGLGSQ-----GTSGR 55

Db 3 GNGAGGSGAPATGGAGGPAGLTVGGAGAGGDSAVAGVIGGAGGAGGAALFAGGA 62

Qy 56 GGLGGGAGAAAAAAGGA-----GCGYGLGSQGTSGRGGLGGCGAGAAAAA 108

Db 63 GGAGSG--GSCAAGGAGGAGGAGGLFASGSGGPGFPASTCTGGAGGTGGAGGLFASGV 121

Qy 109 AAAGGAGCGGCGYGLGSQGTSGRGGLGGCGAGAAAAAAGGAGCGGCGGL--GSQGT 167

Db 122 GGTGGAGSGGTGGVG--GTGAGGLFASGAGGAGSGGTGGAGGTGGAGGLFAGGAG 179

Qy 168 GPGYGPQQGTSGRGGLGG-----GAGAAAAAAGGA--GOGYGGGL--GSQ 215

Db 180 GLG--GGNHFTGGHGGAGSAGLALDGGAGGAGGAATTCTGGAGGAGGKAGLLFSGG 237

Qy 216 TSGRGGLGGQ-----GAGAAAAAAGGAGCGGCGYGLGSQGTSGRGGLGGCGAGA 267

Db 238 AGGSGGAAGTGTGNTSGGAGGAGKAGLLPGSGGAGSGGAGGFGANGSTSCAGAGGA 297

Query Match	30.8%;	Score 1075;	DB 2;	Length 882;
Best Local Similarity	39.9%;	Pred. No. 1.6e-53;		
Matches 301;	Conservative 27;	Mismatches 304;		
Indels 122;	Gaps 32;			

QY 11 GSMASGRGGLGGGAGNAAAAAAAAAGGAGCGGCGGLGSGQTSGRGGLGGGAGAAAAA 70
Db 139 GWWLLNGNGNGSCAPQQTGAGGAAGLLGHGTGGAG--GTGASGKGTGGLWMSGG 196
QY 71 AAAAGGAGCGGCGGLGSGQ-----GTSRGLGG--QGAGAAAAAAAGGA--CGCGYGG 122
Db 197 AGGAGGSG--CGSGAGGNALMFGLGNGGAGGAASGVNGVGGAGGAGGALVAIGGAGG 255
QY 123 LGSQGTSGRGLGGGAGAAAAAAAGGAGCGGCGGLGSGQTSRPGGYC----PQQQT 178
Db 256 AGCAATTTGTCGAGGAGSN-ALGLFLGLGGSGGGSDSAMSGGAGGAGSGGAASPFID 314
QY 179 SGRGGLGGGAGAAAAAAAGAG--QGGYGLGSGQTSRGLGGG 226
Db 315 IGIAGGCHG-GAGTNGGAGGAGGSGTVPFALDLSWGAGNGCAATTGTGGAG--G 371
QY 227 AGAAAAAAAGGAGCGGCGGLG--SQGTSGRGLGGGAGAAAA-----AAAAAGGA 279
Db 372 TGFENAPDFIGFAYGGAGLGGATGAGTGGTGGVGGGFAALGVGVGGAGGAGGA 431
QY 280 -----GQGGYGLGSGQTSGRGLGGGAGAAAAAAAG-----GAGGGYGL 325
Db 432 ATETGGTGGAGGLGVGLGCGAGGAGPG-CAASAGSGHGTGCDALGLGAGTGGVGV 490
QY 326 G-----SQGTSGPGGYPGQQTSGRGLGGGAGAAAAAAAGGAGCGGCGGLGSGQT 380
Db 491 GGAATDTGNGGAGGSGTG-LLGGVGGAGGHHGG--ASVGTGGSGGAGGDFGFGAGGN 547
QY 381 SGRGGLG-----GQAGAAAAAAAGGAGCGGCGGLGSGQTSGRGLGGGAGAAAA 434
Db 548 GGNAGTGVGVNGANGNGSATGALAAVGA-----GAAGDATSGTGGFG--GAGGSAR 600
QY 435 AAAAAAGGAGCGGCGGLGSGQTSGRGLGGGAGAAAAAAAGGAG--QGGYC----- 487
Db 601 GLIFALGGAGAA--GGDASTGVGPGPGPGTGTASSPFGIAIAGGAGAGGAGTSCATG 658
QY 488 -----GLGSGQTSRPGYGPQ-QTSRGLGGGAGAAAAAAAGGAG 532
Db 659 GAGGCGVFEGIAVLGLGFGAAGAGGAATGDGATGGAGFGCAGAGTANFLGFSVLHG-G 717
QY 533 QGGYGLGSGQTSGRGLGGGAGAAAAAAAGGAGCGGCGGLGSGQTSRGLGGG 592
Db 718 AGGAGGTAT-GTGGNGAGGGGGLSPVILIGIGGAGGDDGGALVVG--GMGGDGGDG 774
QY 593 AGAA-----AAAAAGGAGCGGCGGLGSGQTSGRG--GLGCGGAGAAAAAAAGGAG 646
Db 775 GEAVAVGIAVGGAGGAGGAAPTCTCGGAGNGGCDALGLGVGVGNGGNAGTGFGANTGNGG 834
QY 647 QG-----GYGGLGSGQTSRPGYGP 667
Db 835 DTTIVVNGMLAPSTLGYGNGGNGVNGGAG-GTG 867

Search completed: December 18, 2002, 16:22:06
Job time : 29.1216 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 10.4528 Seconds
(without alignments)
2702.183 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHGSMASGRGGL.....GGYGPQQTSGIRRRPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	747	1 SPDL_NEPCL	P19837 nephila cla
2	1277.5	36.6	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	1168	33.4	5263	1 FBOH_BOMMO	P05790 bombyx mori
4	1135	32.5	627	1 SPDL_NEPCL	P46804 nephila cla
5	1106.5	31.7	914	1 WA22_MYCTU	O06794 mycobacteri
6	1044.5	29.9	778	1 Y034_MYCTU	P19333 mycobacteri
7	1028	29.4	957	1 Y278_MYCTU	P56877 mycobacteri
8	988	28.3	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.9	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.9	864	1 ELS_RAT	O99372 rattus norv
11	869	24.9	603	1 YD25_MYCTU	Q10637 mycobacteri
12	853.5	24.4	465	1 GRP2_PHAVU	P10496 phaseolus v
13	775.5	22.2	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.8	641	1 EBN1_EBV	P03211 Epstein-Bar
15	748.5	21.4	1356	1 CA21_ONCMY	O93484 oncorhynch
16	704.5	20.2	750	1 ELS_CHICK	P07916 gallus gall
17	698.5	20.0	1355	1 CA21_RANCA	Q42350 rana catesb
18	691	19.8	730	1 ELS_HUMAN	P15502 homo sapien
19	686	19.6	498	1 Y118_MYCTU	O50615 mycobacteri
20	679	19.4	1733	1 VNUA_PRVKA	P33485 pseudorabie
21	678	19.4	543	1 YP91_MYCTU	O50630 mycobacteri
22	673	19.3	384	1 GRP1_PETHY	P09789 petunia hyb
23	669	19.2	481	1 LOR1_MOUSE	P18185 mus musculu
24	655	18.8	1027	1 CAFF_RIFPA	P30754 riftia pach
25	644	18.4	338	1 GRP_ARATH	P27483 arabidopsis
26	626.5	17.9	1262	1 CA13_CHICK	P12105 gallus gall
27	626	17.9	491	1 YK98_MYCTU	Q10707 mycobacteri
28	626	17.9	515	1 Y140_MYCTU	Q50594 mycobacteri
29	625	17.9	1372	1 CA21_MOUSE	Q01149 mus musculu
30	621	17.8	1466	1 CA13_HUMAN	P02461 homo sapien
31	608	17.4	1372	1 CA21_RAT	P02466 rattus norv
32	608	17.4	1453	1 CA11_CHICK	P02457 gallus gall
33	607.5	17.4	1418	1 CA12_HUMAN	P02458 homo sapien

ALIGNMENTS

RESULT 1

ID	SPDL_NEPCL	STANDARD;	PRT;	747 AA.
AC	P19837;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-!- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			
CC	-!- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M37137; AAA29380.1; -;			
DR	EMBL; U03848; AAB60212.1; -;			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	655	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	25	1.
FT	REPEAT	1	25	2.
FT	REPEAT	26	38	3.
FT	REPEAT	39	66	4.
FT	REPEAT	67	96	5.
FT	REPEAT	97	130	6.
FT	REPEAT	131	158	

34	604.5	17.3	1460	1	CA11_CANFA	Q9xs17	canis famil
35	603.5	17.3	1049	1	CA13_BOVIN	P04258	bos taurus
36	602.5	17.2	463	1	YA68_MYCTU	O53416	mycobacteri
37	602	17.2	1464	1	CA11_HUMAN	P08123	homo sapien
38	598	17.1	1366	1	CA21_MOUSE	P28481	mus musculu
39	598	17.1	1459	1	CA12_MOUSE	P08121	mus musculu
40	597.5	17.1	1464	1	CA21_MOUSE	O46392	canis famil
41	597	17.1	1366	1	CA21_BOVIN	P02465	bos taurus
42	595	17.0	1364	1	CA11_MOUSE	P11087	mus musculu
43	594.5	17.0	1453	1	CA21_CHICK	P02467	gallus gall
44	573.5	16.4	1362	1	CA21_CHICK	P17140	caenorhabdi
45	563	16.1	1758	1	CA24_CAEEL		

FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 682 V -> L (IN REF. 1).
FT CONFLICT 672 S -> T (IN REF. 1).
FT CONFLICT 695 NFGSGCDVLIQALLVWSALIQILGSSSIGOVNYSAGQA
TOIGQSVYQAL -> ILVFLDVMSSFKLFSRLFLLSRS
(IN REF. 1).
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;
Query Match 61.1%; Score 2135; DB 1; Length 747;
Best Local Similarity 68.3%; Pred. No. 7.9e-89;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;
QY 23 QGAGAAAAAAGAGAGAGGGYGLGSGQT--SGRGGLGGGAG-AAAAAAGAGAGG 79
DB 1 QGAG-----AAAAAAGAGAGGGYGLGSGAGGGYGLGSGAGAGAGAAAAAGAGG 56
QY 80 GYGGLGSGQTSRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSRGGLGGGAGA 139
DB 57 GYGGLGSGQ-----AGRGQGAG-----AAAAAAGAGAGGGYGLGSGQ-AGRGGLGGGAGA 107
QY 140 GAAAAAAGAGAGAGGGYGLGSGQTSRGGYGPQQTSGRGGLGGGAGAGAAAAAAG 199
DB 108 G-----AAAAAAGAGAGGGYGLGSGQ-----AGRGQGQ-----AAAAA 142
QY 200 AGAGAGGGYGLGSGQTSRGGLGGGAGAGAAAAAAG-----AGAGAGGGYGLG 250
DB 143 AGAGAGGGYGLGSGQ-AGRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGAGAGGGYGLG 201
QY 251 SQGTSRGGLGGGAGAGAAAA-----AAAAAAGAGAGGGYGLGSG 289
DB 202 SQG-AGRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGAGAGAGAGAGGGYGLGSG 260
QY 290 QGTSRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSRGGYGP-GQQTSGRGGL 348
DB 261 QG-----AGRGGEGAG-----AAAAAAGAGAGGGYGLGSGQ-AGQGGYGLGSGAGRGGL 311
QY 349 GGGAGAGAA-----AAAAAAGAGAGGGYGLGSGQTSRGGLGGG 390
DB 312 GGGAGAGAAAGAGAGGGYGLGSGAGAGAGAGAGAGAGGGYGLGSGQ-AGRGGLGGG 370
QY 391 AGAAAAAAGAGAGGGYGLGSGQTSRGGLGGGAGAGAAAAAAGAGAGAGGGYGLG 450
DB 371 AG-----AAAAAAGAGAGGGYGLGSGQ-----AGRGQGAG-----AAAAAAGAGAGGGYGLG 419
QY 451 LGSQTSRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSRGGLGGG-----SGP 497
DB 420 LGNQG-AGRGGLGGGAG--AAAAAAGAGAGGGYGLGSGQTSRGGLGGG-----SGP 475
QY 498 GYGYP-GQQTSGRGGLGGGAGAGAAAAAAG-----AGAGAGGGYGLGSGQTSRGGLG 550
DB 476 GYGGLGSGQAGR--GGGAGAGAGAGAGAGGGYGLGSGQTSRGGLGGG-----SGP 531
QY 551 GGGAGAGAA-----AAAAAAGAGAGGGYGLGSGQTSRGGLGGG-----SGP 590
DB 532 GGGAGAGAGAGAGGGYGLGSGQTSRGGLGGG-----AGRGG 587

QY 591 QGAGAAAAAAGAGAGGGYGLGSGQTSRGGLGGGAGAGAAAAAAGAGAGGGY 650
DB 588 QGAG-----AAAAAAGAGAGGGYGLGSGQV-GRGGLGGGAG-----AAAAAAGAGAGGGY 636
QY 651 GGLGS 655
DB 637 GVGGS 641
RESULT 2
YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL022022; CAAL17745.1; -.
DR TubercuList; RV3508; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00534; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 36.6%; Score 1277.5; DB 1; Length 1901;
Best Local Similarity 39.7%; Pred. No. 1.1e-50;
Matches 351; Conservative 28; Mismatches 270; Indels 235; Gaps 39;
QY 11 GSWASRGGLGGQ-GAGAAAAAAGAGAGGGYGLGSGQ-----GTS 53
DB 292 GGALGGTGTGTGGAGAGGRCALLLGGAGGGGGLGGAGGGGTGGAGDGLVGGVGGTG 351
QY 54 GRGGLGG-----QGAGAAA-----AAAAAAGG-AGGGYGLGSGQ----- 89
DB 352 GKGGGVAGLGGAGAGACQLFSAGGAGAVGGVGGGGAGGAGAGADAPASTGLRG 411
QY 90 ----TSRGGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSR-----GGLGG 136


```

QY 593 -----AGAAAAAAGAGAGGGG-LGSGTSGRGLGGGAGAGAAAAAAGAGAG 647
Db 822 RSGFGAGGAGGASDTGACGNGCAGGAGKAGLYGNGGAGGAGGAGTSGKGGAGNAVIGN 881

QY 648 GGVGGLGSG--GTSRPGGYG 665
Db 882 GNGGNAGAGGTAGAGGAG 901

RESULT 6
YQ34_MYCTU
ID YQ34_MYCTU STANDARD; PRT; 778 AA.
AC P71933;
DT 01-NOV-1997 (Rel. 35, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical PE-PGRS family protein rv2634c.
GN RV2634C OR MT2712 OR MTCY441.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland L., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam J.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z80225; CAB02341.1; .
DR EMBL; AF007103; AAK47026.1; ALT_INIT.
DR TIGR; MT212; -.
DR Tuberculist; RV2634c; .
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein: Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 29.9%; Score 1044.5; DB 1; Length 778;

```

```

Best Local Similarity 41.3%; Pred. No. 1.4e-40;
Matches 289; Conservative 21; Mismatches 290; Indels 99; Gaps 34;

QY 11 GSNASGRGGLGGGACAAAAAAGAGGAGGCGGGLGSG-----GTSRGG--LG 59
Db 135 GGLLNGNGGSGAPQPGCGAGDAGLNGGTGGGDLVGSAAAGVGRGGLLG 194

QY 60 GCGAGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGAAAAAAGAGAGG 119
Db 195 NGGTGAGGAAGATLVG--GTGGVG--GATCLIGSGFG--GAGAAAGVCTTGGVSGG 249

QY 120 YGGLGSGTSGRGLGGGAGAAAAAAGAGAGGCGGGLGSGTSGPGGPGGQTS 179
Db 250 VGSV--FGNGGFGAGGLGAAGGVGAASYFCTGGGGVGGDGAPG--GDGGAGP--LLI 303

QY 180 GRGGLGGGAGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGAAAAAAG 239
Db 304 GNGVGGLG--GAGAA-----GNGGAGGM-----LLDGLGAGGCGGFAVGLGMPG 350

QY 240 GAGQG-----GYGGLGSGTSGRGLGGGAGAAAAAAG-----GAGQGCGGL 287
Db 351 AGNGGNANWFGSGAGGCGGTCLACTNGVNPCSIANPTGANGTDSNGNGQTGNGGP 410

QY 288 GSGTSGR--GGLGGGAGAAAAAAGAGGCGGGLGSGTSGPGGPGGQTSGRG 346
Db 411 GPAGGVGEAGVGCGQ-----GLGESLDGNDGTGKGGAG--GTAGTDG--GAG 458

QY 347 GLGGGAGAAAAAAGAGGCGGGLGSGTSGRGLGGGAGAAAAAAGAGAG 406
Db 459 GAGGIGETGSGAGVATGEGDGTGVDG--GVGAGGKGGGHTGVGDADFSGDGLG 517

QY 407 QGGYGLGSGTSGRGLGGGAGAAAAAAGAG--GQGGYGLGSGTSG--RGGLG 464
Db 518 GDENGALGAAG--GNGGTGAGNGGRCGMLCNGCAGGAGGTGCTGGGGAAGFAGVG 575

QY 465 QGAGAAAAAAGAGGCGGGLGSGTSGPGGPGGQTSGRGLGGGAGAAAAA 524
Db 576 AGGEGLTGAGTAEGGT--GGLGLG--GVGTTGGMG-----GSGGVGNGAGSLIG 625

QY 525 AAAAGGA--CGGCGGLGSGTSGRGLGGG--GAGAAAAAAGAGGCGGGLGSGT 582
Db 626 LGGGGAGGCGGTGGIG--GIGAGNGGAGGAGTTGGGATIGGGGGTGGVGGAG--GT 681

QY 583 SGRGGLGG--GAGAAA-----AAAAAAGAGGCGGGLGSGTSGRGLGGGAGAAA 635
Db 682 GCTGGAGTTGSGGAGGLGWAGAGGTGAGTGGCGGLGGG--GNGSGGTGATGG 739

QY 635 AAAAAGAGGCGGCG-----LGSGTSGPGGPGGCG 669
Db 740 GGDFA LGNGGAGGAGGSPGSGGIGGNMGP-----PGTQ 774

RESULT 7
Y278_MYCTU
ID Y278_MYCTU STANDARD; PRT; 957 AA.
AC P56877;
DT 30-MAY-2000 (Rel. 39, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical PE-PGRS family protein rv0278c precursor.
GN RV0278C OR MT0291 OR MTV035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Davies R., Davies R., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AL021958; CAA17514.1; -
EMBL; AE006968; AAK45011.1; -
TIGR; MT0772.5; -
TubercuList; RV0747; -
InterPro; IPR000084; PE_region.
Pfam; PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL
FT CHAIN 31 801 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV0747.
FT CONFLICT 188 188 R -> S (IN REF. 2).
FT CONFLICT 225 225 R -> G (IN REF. 2).
FT CONFLICT 227 227 R -> G (IN REF. 2).
FT CONFLICT 295 295 K -> R (IN REF. 2).
FT CONFLICT 300 300 S -> G (IN REF. 2).
FT CONFLICT 300 300 S -> I (IN REF. 2).
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 377 377 A -> P (IN REF. 2).
FT CONFLICT 577 577 T -> A (IN REF. 2).
FT CONFLICT 580 595 MISSING (IN REF. 2).
SQ SEQUENCE 801 AA; 65407 MW; EA54C9BF45A00F41 CRC64;

Query Match 28.3%; Score 988; DB 1; Length 801;
Best Local Similarity 37.9%; Pred. No. 4.4e-38;
Matches 284; Conservative 33; Mismatches 289; Indels 144; Gaps 31;

QY 22 GCGAGAAAAA-----AAAGGAGCGGCGLGSQCTSGRGGLGQ--GAGAAA 67
DB 83 GGGAYAAEAAYVSLPLAINAOFVATGRPLIGNGANGAPGTGGNGPGWILGNAGG 142
QY 68 AAAAAAGAGGAGGCGGL--GSGQTSG-----RGGLGCGAGAAAAAAGAGGCGG 121
DB 143 GSGAPGAGGNGGAGGLFCGSGAGGASTDVAGAG--GAGGAGGNAGMLFGAGVGGVG 200
QY 122 GLGSGQ--TSRGGLGGGAGAAAAAAGAGAGQ--GGYGLG-----SQGTSPPGGY 172
DB 201 GFSNGCATGTGAGGAGGAG--GLFGAGRGRGSGSGNLTGGAGGAGNAGTLATDGGAGGT 259
QY 173 GPCQOTSRRGLGCGAGAAAAAAGAG--GAGGAGGAGGAGGAGGAGGAGGAGGAGG 218
DB 260 GGAASRGFGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 319
QY 219 RGGI--GGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 265
DB 320 NGGASTGGDGGPGGAGGTGVLIGNGNGSGGTGA--TLGKAGIGGTGGVLLGLDGF 377
QY 266 -----GAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290
DB 378 PASTSPLHTLQDDVINMVPQTLTGRPLIGNANGCTPGTGADGGAGGWLFCNGNGGQ 437
QY 291 GTSRGGLGCGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
DB 438 GTI--GGVNG--GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 494
QY 351 QGA-----GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 403
DB 495 SGAGAVGGNGGAGGNAGLLGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551
QY 404 G----AGQGGYGLGSG--GTSRGGGLGGCGAGAAAAAAGAGGAGGAGGAGGAGGAG 458
DB 552 GFGSPAGAGGIGGAGGNGGLFGAGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 601
QY 459 RGLGCGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 518
DB 602 AGGTGAGGSHSTAAGYSSGAGGA--GGDAGLLSLGAGGAGGAGGAGGAGGAGGAGGAG 657

Db 185 VYPQVAGIGAGGKPKVPYVGLPGVLPCTGAREPFGVGLPGVPTGTGVKAKAPG 244
Qy 150 AGGACQG--GYGGLGSQ-----GTSGPGGY-----PGQOTSGRGLGGQ----- 187
Db 245 GGGAFSGIPGVPFGGQPGVPLGYPIKAPKLPKPGYGLPTNGKLPYGVAGAGKAGYPT 304
Qy 188 --GAGAAAAAAGAGAGAGG--GLSQGTSGR-----GGLGGGAGAAAAA 235
Db 305 GTGVGSQAAAAAANKYAGAGAGVLPVCGGGGIPGGAGATPGIGIAGAGTPAAAAA 364
Qy 236 AAAGCA-----GOGGYGLGSQGTSGRGLGGGAG 266
Db 365 AAANKAYGAAGLVPGPGVRLPGAGIPGVGIPGVGPGIGGPGIYGGPGAV 424
Qy 267 AAAAAAAGCA--GOGGYG--GLSQGTSGRGLGGGAGAAAAAAGGAGQ 319
Db 425 SPAAAAAANKAANKYARGGVGITYGVGAGGFPYGVGAGAGLGAGGASPAAAAAA 484
Qy 320 GYGGLGSQGTSGR--GYGPG--QOTSGRGLGGGAGAAAAA-----AG-CAG 368
Db 485 YGAGAGALGLVPGAVPGALPGAVPAGVPGAGVPGAGTPAAAAAANKAANKAGLGP 544
Qy 369 QGGY--GGLGSQGTSGRGLGGGAGAAAAAAGAGGAGGAGGGLGSQGTSGRGLGGQ 427
Db 545 VGGVPGVGVGIPGVGVG--VPGGVPGVGTGIG--AGPGGLGA 588
Qy 428 GAGAAAAA-----AAAG--GAGGGY--GLGSQGTSGRGLGGGAGAAAAA 476
Db 589 GSPAAKSAANKAANKAYRAAAGLGAAGVPGFAGAGVPGFAGAGVPGFAGAGVPGF 648
Qy 477 AAG--GAG-----GOGYGLGSQ--GTSGPGYGPQGTSGRGLGGQ-- 516
Db 649 GAGVPGFAGVPGVPSLAASKAANKYGAAGLGGPGGLGGPGLG-----GPGGLGGAGVP 702
Qy 517 ---AGAA--AAAAAAGAGAGGAGGGLGSQGTSGRGLGGGAGAAAAAAGAGGAGQ 572
Db 703 GRVAGAAPAAAAAANKAANKAYGLGAGGLGAGGLG-----AGGLGAG 753
Qy 573 GY--GGLGSQGTSGRGLG--CGAGAAAAAAGAGGAGGAGGGLGSQGTSGRGLGGQ 629
Db 754 GLGAGGLGAGGL--GAGGLGAGGGSVPAANKAANKAYGAAGLGGV--LGARPPFGGVAAP 810
Qy 630 GAGAAAAAAGAGAGG--YGG--LGSQGTSGPGGYG 665
Db 811 GFLSPYIPGGAGGLGVGGKPPKPYGALGALGTQGGGCGF 852

RESULT 10

ID	ELS_RAT	STANDARD	PRT	864 AA.
AC	Q99372			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Elastin precursor (Tropoelastin) (Fragment).			
GN	ELN.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RX	MEDLINE=91104868; PubMed=1702999;			
RA	Pierce R.A., Deak S.B., Stollie C.A., Boyd C.D.;			
RT	"Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";			
RL	Biochemistry 29:9677-9683(1990).			
RN	[2]			
RP	SEQUENCE OF 781-864 FROM N.A.			
RX	MEDLINE=88330868; PubMed=2971041;			
RA	Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;			
RT	"Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";			
RL	J. Biol. Chem. 263:13504-13507(1988).			

RN [3]
RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=92241859; PubMed=1572837;
RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;RT "Elements of the rat tropoelastin gene associated with alternative
splicing.";
RL Genomics 12:651-658(1992).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: Various isoforms are produced by alternative
splicing.
CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M60647; AAA42269.1; -
DR EMBL; J04035; AAA42268.1; -
DR EMBL; M86372; AAA42271.1; -
DR EMBL; M86355; AAA42271.1; JOINED.
DR EMBL; M86363; AAA42271.1; JOINED.
DR EMBL; M86364; AAA42271.1; JOINED.
DR EMBL; M86366; AAA42271.1; JOINED.
DR EMBL; M86371; AAA42271.1; JOINED.
DR EMBL; M86376; AAA42272.1; -
DR EMBL; M86373; AAA42272.1; JOINED.
DR EMBL; M86375; AAA42272.1; JOINED.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 864 ELASTIN.
FT VARSPIC 263 307 MISSING (IN CERTAIN ISOFORMS).
FT VARSPIC 308 308 MISSING (IN CERTAIN ISOFORMS).
FT VARSPIC 809 823 MISSING (IN CERTAIN ISOFORMS).
SQ SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;
Query Match 26.9%; Score 939; DB 1; Length 864;
Best Local Similarity 38.3%; Pred. No. 6.8e-36;
Matches 316; Conservative 32; Mismatches 279; Indels 198; Gaps 38;
Qy 18 GGLGGGAGAAAAAAG-----GAGGGYGGGLGSQ-----TSGRGG--LGGGGAGAA 66
Db 53 GGLGGGALPGGKPKPGAGLLGAFAGPGGLGAGPGAGLSYASRPGVLPVGGGAGAA 112
Qy 67 AAAAAAAGAGGAGGAGGGL-----GSGG-----TSGRGLGGGAGAGAAAAA 109
Db 113 AAYKAAKAGAGLGGIGGPGVGVGAVGVGVGAVGVGVGAVGVGVGAVGVGVGAVGV 172
Qy 110 AAAG--GAG--OGGYGLGSQGTSGRGLGGGCA-----GAAAAAAGGAA 153
Db 173 LGAGVGGGKPGKPGVGLPGVPGVLPGTGTPGVPVGTGTGKAKVPGGGG 232
Qy 154 GGGYGGGLGSQGTSG-----PGGYG-----PGQOTSGRGLGGQ-----G 188
Db 233 AFSGIPGVGPGGQPGVPLGYPTKAPKLPGGYGLPYTNGKLPYGVAGAGKAGYPTGTG 292
Qy 189 AGAAAAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 237
Db 293 VGSQNAVAANKAANKYAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 352
Qy 238 AGGAGGGYGGGLGSQGTSGR--GGLGGGAGAGAAAAAAG--ACGGYGGGLGSQGTSGR 295
Db 353 AKAAYGAAGGLVPGGPGVVPVAGIPGVGIPGVGIPGVGIPGVGIPGVGIPGVGIPGVGIP 412

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 39,982 Seconds
(without alignments)
3509.535 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHSWASRGGL.....GCVGPQQQTSGIRPPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2009	57.5	644	5 Q8WSW4	Q8WSW4 nephila cla
2	1991.5	57.0	617	5 Q46172	Q46172 nephila cla
3	1726.5	49.4	544	5 Q46171	Q46171 nephila cla
4	1592	45.6	691	5 Q9BIU3	Q9BIU3 dolomedes t
5	1568	44.9	648	5 Q9BIU7	Q9BIU7 argiope tri
6	1531.5	43.8	854	5 Q9BIU4	Q9BIU4 dolomedes t
7	1450	41.5	988	5 Q17434	Q17434 nephila cla
8	1448	41.5	1953	5 Q9BIT7	Q9BIT7 nephila cla
9	1386.5	39.7	2639	5 Q76786	Q76786 antheraea p
10	1350	38.6	2655	5 Q964F4	Q964F4 antheraea p
11	1267	36.3	1217	16 Q8VIY9	Q8VIY9 mycobacteri
12	1263.5	36.2	1715	16 Q8VI20	Q8VI20 mycobacteri
13	1254.5	35.9	1489	16 Q53559	Q53559 mycobacteri
14	1228.5	35.2	1079	16 Q53557	Q53557 mycobacteri
15	1199.5	34.3	447	5 Q9BIV1	Q9BIV1 argiope aur
16	1197	34.3	1306	16 Q53775	Q53775 mycobacteri

17	1186.5	34.0	1384	16 Q8VI21	Q8VI21 mycobacteri
18	1182.5	33.9	1381	16 Q53552	Q53552 mycobacteri
19	1173	33.6	1685	16 Q53215	Q53215 mycobacteri
20	1160.5	33.2	1507	16 Q8WJ23	Q8WJ23 mycobacteri
21	1159.5	33.2	1884	5 Q9NHW2	Q9NHW2 nephila mad
22	1158	33.2	1002	5 Q9BIU8	Q9BIU8 argiope tri
23	1157	33.1	738	5 Q02402	Q02402 pinctada fu
24	1157	33.1	1538	16 Q53395	Q53395 mycobacteri
25	1152.5	33.0	1408	16 Q8VK17	Q8VK17 mycobacteri
26	1151	33.0	2249	5 Q9NHW4	Q9NHW4 nephila cla
27	1148	32.9	1468	5 Q9NHW2	Q9NHW2 nephila me
28	1145	32.8	1329	16 Q06810	Q06810 mycobacteri
29	1139	32.6	853	16 Q53439	Q53439 mycobacteri
30	1114.5	31.9	749	16 Q53844	Q53844 mycobacteri
31	1099	31.5	907	5 Q44359	Q44359 mycobacteri
32	1083	31.0	871	5 Q44358	Q44358 nephila cla
33	1080.5	30.9	879	16 Q8VKD2	Q8VKD2 mycobacteri
34	1075	30.8	882	16 Q53845	Q53845 mycobacteri
35	1041.5	29.8	563	5 Q9BIT5	Q9BIT5 nephila mad
36	1036	29.7	767	16 Q53435	Q53435 mycobacteri
37	1024	29.3	837	16 Q53684	Q53684 mycobacteri
38	1015.5	29.1	651	5 Q9BIU9	Q9BIU9 argiope tri
39	1014	29.0	783	16 Q53809	Q53809 mycobacteri
40	1011	28.9	636	5 Q16987	Q16987 araneus dia
41	1005	28.8	775	16 Q8VJ15	Q8VJ15 mycobacteri
42	1004	28.7	731	16 Q50415	Q50415 mycobacteri
43	984.5	28.2	714	16 Q53556	Q53556 mycobacteri
44	984	28.2	741	16 Q06808	Q06808 mycobacteri
45	975	27.9	810	11 Q06E29	Q06E29 mus musculu

ALIGNMENTS

RESULT 1

Q8WSW4 ID Q8WSW4 PRELIMINARY; PRT; 644 AA.
AC Q8WSW4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Dragline silk protein (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RA Basel R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061814; AAL32375.1; -;
DR InterPro; IPR00104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
FT NON_TER 644
SQ SEQUENCE 644 AA; 52104 MW; 9024AF2F65C5F75A CRC64;

Query Match	57.5%	Score	2009	DB 5	Length	644
Best Local Similarity	70.9%	Pred. No.	7.9e-101			
Matches	461	Conservative	12	Mismatches	39	Indels 138; Gaps 30
QY	30	AAAAAAGGAGGGYGGGLGSGQTSRGGGLGGGAGAAAAAAGGAGGGYGGGLGSGQ	89			
Db	2	AAAAAAGGAGGGYGGGLGSGQ-----AGRGQGGAG-----AAAAAAGGAGGGYGGGLGSGQ	53			
QY	90	TSRGGGLGGGAGAGAAAA-----AAAAAAGGAGGGYGGGLGSGQT	128			
Db	54	AGRGGLGGGAGAGAAAAAAGGVCQGGGLGGGAGAGAAAAAAGGAGGGYGGGLGSGQ-	111			
QY	129	SRGGGLGGGAGAGAAAAAAGGAGGGYGGGLGSGQTSRGGYGP-QQTSRGGGLGGQ	187			
Db	112	---AGRGGAGAGAAAAA---GAGGGYGGQ---AGGGYGGGLGSGQAGRGGGLGGQ	160			
QY	188	GAGAAAAAAGGAGGGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGGYGGY	247			

Db	161	GAG---AAAAAAGGAGQ-----GGUGGQAGQGACAAAAGAGAGQGGYG	203
Qy	248	GLGSQGTSGRGLGGCGAGAAAAAAAGGAGQGGYGGGLSQGTSGRGLGGCGAGAAA	307
Db	204	GLGSQ-ACRGSGCGAG-----AAAAAAGGAGQGGYGGGLSQG-ACRGLGGQAG---	254
Qy	308	AAAAAAGAGQGGYGGGLSQGTSPPGYGP-GQOTSGRGLGGQAGAAAAAAGG	366
Db	255	AAAAAAGAGQGGYGGGLSQG-AGQGGYGGGLSQGACRGLGGQAG---AAAAAAGG	310
Qy	367	AGGGYGGGLSQGTSRCRGLGGQAGAAAAAAGAGAGQGGYGGGLSQGTSRCRGLGG	426
Db	311	AGQ-----GGUGGQAGQGACAAAAGAGQGGYGGGLNQ-ACRGQGC-	354
Qy	427	QGAGAAAAAAGAGQGGYGGGLSQGTSRCRGLGGQAGAAAAAAGAGAGQGGY	486
Db	355	-----AAAAAAGAGQGGYGGGLSQG-ACRGLGGQAG-----AAAAAAGAGQGGY	401
Qy	487	GGLSQGTSGPGYGP-GQOTSGRGLGGQAGAAAAAAGAGAGQGGYGGILGSGT	545
Db	402	GGLGQ-AGQGGYGGGLASGGSRGLGGQAG-----AAAAAAGAGQ-	444
Qy	546	GRGLGGQAGAAAAAAGAGAGQGGYGGILGSGTSCRGLGGQAGAAAAAAGG	605
Db	445	--GGLGQAGQAGAAAAAAGGVQGGYGGILGSGQ----ACRGQGGAG----	494
Qy	606	GAGQGGYGGILGSGTSGRGLGGQAGAAAAAAGAGAGQGGYGGILGSGS	655
Db	495	GAGQGGYGGILGSGV--GRGLGGQAG-----AAAAAAGAGQGGYGGVVS	537
RESULT 2			
O46172	O46172	PRELIMINARY;	PRT; 617 AA.
ID	O46172		
DT	O46172:		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Drighline silk protein spidroin 1 (Fragment).		
OS	Nephila clavipes (Orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.		
OX	NCBI_Taxid=6915;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RA	MEDLINE=94165058; PubMed=8120021;		
RX	Beckwitt R., Arcidiacono S;		
RT	"Sequence conservation in the C-terminal region of spider silk		
RT	proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus		
RT	bicentariarius (Araneidae).";		
RL	J. Biol. Chem. 269:6661-6663(1994).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RA	MEDLINE=98318805;		
RX	Beckwitt R., Arcidiacono S., Stote R.;		
RT	"Evolution of repetitive proteins: spider silks from Nephila clavipes		
RT	(Tetragnathidae) and Araneus bicentariarius (Araneidae).";		
DL	Insect Biochem. Mol. Biol. 28:121-130(1998).		
DR	EMBL; U37520; AAC04504.1; -		
FT	NON_TER	1	
FT	NON_TER	617	
SQ	SEQUENCE	617 AA; 49410 MW; 1EF0CE9269A832E2 CRC64;	

Db	63	AAAAAAAAAGGVGQ-----GGGLGGGAGOGAGAAAAAAGACGGYGGGLGSQG	108
Qy	128	TSGRGLGGGOGAGAAAAAAGGAGOGGYGGLGSQGTSGPGYGPGQOTSRRGSLGSG	187
Db	109	-AGRGGSGGQAG---AAAAAAGGAGOGGYGGLGSQG-----AGRGGGLGQ	150
Qy	188	GAGAAAAAAAAG--GAGOGGYGGLGSQGTSGRGLGGOGAGAAAAAAGAA	237
Db	151	GAGAAAAAAGGAGOGGYGGLGGGAGOGGYGGLGSQG-AGRGLLGQAG--AAAAA	206
Qy	238	AGGAGOGGYGGLGSQGTSGRGLGGOGAGAAAAAAGGAGOGGYGGLGSQGTSGRG	297
Db	207	AGGAGQ-----GGTGGGAGOGAGAAAAAAGGAGOGGYGGLGSQG-----AG	248
Qy	298	LGGOGAGAAAAAAGGAGOGGYGGLGSQGTSGPGGYCP-CQOTTSRRGLGGOGAGA	356
Db	249	RGGOGAGAAAAAV----GAGOGGYGGQ-----AGOGGYGGLGSQAGRGGLGGOGAG--	298
Qy	357	AAAAAAGGAGOGGYGGLGSQGTSGRGLGGOGAGAAAAAAGGAGOGGYGGLGSQ	416
Db	299	-AAAAAAGGAGQ-----GGTGGGAGOGAGAAAAAAGGAGOGGYGGLGNQ	343
Qy	417	GTSRRGLGGOGAGAAAAAAGGAGOGGYGGLGSQGTSGRGLGGOGAGAAAAAAGAA	476
Db	344	G-AGRGGQ-----AAAAAAGGAGOGGYGGLGSQG-AGRGLLGQAG-----AAAA	388
Qy	477	AAGGAGOGGYGGLGSQGTSGPGGYCP-CQOTTSRRGLGGOGAGAAAAAAGGAGOGG	535
Db	389	AAGGAGOGGYGGLGGQG-AQOGGYGGLGSQSGSRGLGGOGAG-----AAAAAGGAGQ--	441
Qy	536	YGGLGSQGTSGRGLGGOGAGAAAAAAGGAGOGGYGGLGSQGTSGRGLGGOGAGA	595
Db	442	-----GGTGGGAGOGAGAAAAAAGGVRGGYGGGLGSQ-----AGRGGGAG--	484
Qy	596	AAAAAAGGAGOGGYGGLGSQGTSGRGLGGOGAGAAAAAAGGAGOGGYGGLGS	655
Db	485	-----AAAAAAGGAGOGGYGGLGGQV-GRGLGGOGAG-----AAAAAGGAGOGGYGVS	534
 RESULT 3 O46171 PRELIMINARY; PRT; 544 AA.			
ID	O46171;		
AC	01-JUN-1998 (TrEMBLrel. 06, Created)		
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Spidroin 1 (Fragment).		
OS	Nephila clavipes (Orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
CC	Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.		
OX	NCBI_TaxID=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=90384959; PubMed=2402494;		
RX	Xu M., Lewis R.V.;		
RT	"Structure of a protein superfiber: spider dragline silk.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).		
RN	[2]		
RP	SEQUENCE OF 449-544 FROM N.A.		
RX	MEDLINE=98148687; PubMed=9487707;		
RA	Arcidiacono S., Mello C., Kaplan D., Cheley S., Bayley H.;		
RT	"Purification and characterization of recombinant spider silk		
RT	expressed in Escherichia coli.";		
RL	Appl. Microbiol. Biotechnol. 49:31-38(1998).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98318805;		
RA	Beckwith R., Arcidiacono S., Stote R.;		
RT	"Evolution of repetitive proteins: spider silks from Nephila clavipes		
RT	(Tetragnathidae) and Araneus bicentarius (Araneidae).";		
RL	Insect Biochem. Mol. Biol. 28:121-130(1998).		
RN	[4]		

Db 959 GPGGYPGQGGPCGAAAAAGSGPGYGP-GQGGPGGCGAAAAAGRGGPGYGPQGGGPG 1017
QY 267 AAAAAAAGAGAGGGY-----GLGSGTSGRG-----GLGQGAG 304
Db 1018 GPCAAAAAA-CRGGPGYGPQGGPGGCGAAAAAGRGGPGYGPQGGPGGAA 1076
QY 305 AAA-----AAAAAGAGGGY-----GLGSGT-----GPGYGP 339
Db 1077 AAAGPGYGPQGGPCGAAAAAGRGGPGYGPQGGPGGCGAAAAAGRGGPGYGP 1136
QY 340 QGTSGRGLGGGAGAAAA-----AAAAAGAGGGYGLGSGTSGRG 386
Db 1137 QGGPGGPGGCGAAAAAGRGGPGYGPQGGTGAAGAAAAAGAGGAGY-----PGQGP 1189
QY 387 GGGGAGAAAA-----AAAAAGAGGGY-----G 412
Db 1190 GCGPAAAAAGPGYGPQGGPCGAAAAAGSGPGYGPQGGPGGSSAAAAAGPGRYG 1249
QY 413 LGSQGT-----SGRGL--GGGAGAAAAAAGAGAGGGYGLGSGTSGRG- 460
Db 1250 PGQGGPCGAAAAAGSGPGYGPQGGPGGCGAAAAAGRGGPGYGP-GQGGPGGPGA 1308
QY 461 -----GLGGGAGAAAA-----AAAAAGAGGGY-- 487
Db 1309 AAAAGPGYGPQGGPCGAAAAAGSGPGYGPQGGPGCGAAAAAGRGGPGYGG 1368
QY 488 --GLSQ-----TSRGGYGPQ-----TSRGL--GGGAGAAAAAAG 529
Db 1369 QGGPGGAAAAAGPGYGPQGGPCGAAAAAGSGPGYGPQGGPCRSAAAAA 1428
QY 530 GAGGGY-----GLSQ-----TSRG-----GLGG 552
Db 1429 GRGGYGPQGGPGGAAAAAGPGYGPQGGPCGAAAAAGRGGPGYGPQGGPGG 1488
QY 553 QGAGAAAA-----AAAAAGAGGGYGLGSGTSGRG----- 586
Db 1489 SGAAAAAGRGGYGPQGGPGCGAAAAAGTGGYGP-GQGGPGCGAAAAAGR 1547
QY 587 -----GLGGGAGAAAA-----AAAAAGAGGGYGLGSGTSGR 623
Db 1548 GPGYGPQGGPGGCGAAAAAGPGYGPQGGPCGAAAAAGSGPGYGP-GQGGPGG 1606
QY 624 G-----GLGGGAGAAAA-----AAAAAGGACQ 647
Db 1607 GAAAAAGRGGYGPQGGPGGAAAAAGRGGPGYGPQGGPGGSGNAAGSGRGP 1666
QY 648 GGYG-----GLSQ-----TSRGGYGPQGTSGIRPAA 678
Db 1667 GYGPQGGPGGCGAAAAAGPGYGPQGGPCGAAAAA 1705
RESULT 9
ID 076786 PRELIMINARY; PRT; 2639 AA.
AC 076786;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Fibroin.
OS Antherea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Saturniini; Antherea.
OX NCBI_TaxID=71119;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20496953; PubMed=11040284;
RA Sezutsu H., Yukihiko K.;
RT "Dynamic rearrangement within the antherea pernyi silk fibroin gene
is associated with four types of repetitive units.";
RL J. Mol. Evol. 51:329-338(2000).
DR EMBL: AF083334; AAC32605.1;
SQ SEQUENCE 2639 AA; 216057 MW; 2EE3310DEEB09B9A CRC64;

Query Match 39.7%; Score 1386.5; DB 5; Length 2639;
Best Local Similarity 40.4%; Pred. No. 5e-67;
Matches 402; Conservative 23; Mismatches 239; Indels 330; Gaps 34;
QY 2 RGSHHHHHGSMA-----SGRGLGG-----QGAGAAAAAAGAGGGY 45
Db 135 RGGGYETHNSYSSYSGSSAAAGAGGGYGGYCGSDSAAAAAAGAGGGY 194
QY 46 GLCSQ-----GTSRGLGGGAGAAAAAAGAGGGYGGY 86
Db 195 GYGSDSAAAAAAGAGAGGGYGGYCGSDSAAAAAAGAGAGAG 254
QY 87 SQGTSGRGLGGGAGAAAAAAGAGGGYGGYCGSDS----- 129
Db 255 GYGWGDGGYGSAAAAAAGAGAGGGYGGYCGSDSAAAAAAGAGSS 314
QY 130 -----GRGGLGGGAGAAAAAAGAGGGYGGYCGSDS----- 164
Db 315 AGGAGGGYCGDGGYCGSDSAAAAAAGAGAGGGYGGYCGSDSAAAAA 374
QY 165 -----GTSGPGYGPQGTSGRGLGGGAGAAAAAAGAG 201
Db 375 ARRACHDRAGSAAAAAAGAGAGGGYGGYCGSDSAAAAAAGAGAG 433
QY 202 GAGGGYGG-----LGSQGTSGRGLGGGAGAAAAA 235
Db 434 GSGAGAGGGYCGDGGYGSAAAAAAGAGAGGGYGGYCGSDSAAAAA 493
QY 236 AAAGAGGGYGLGSGTSGRGLGGGAGAAAAAAGAGGGYGGYCGSD 291
Db 494 AAAGAGAGGAG--GSYGW-GDGGYCGSDSAAAAAAGAGAGGGY 550
QY 292 T-----SGRGLGGGAGAAAAAAGAGAG--GGY-----GGLGSG 330
Db 551 AAAAAAASARRACHDRAGSAAAAAAGAGAGGGYCGDGGYCGSDSAAA 610
QY 331 -----SGPGYGPQGTSGRGLGGGAGAAAAAAGAGAG--GGAGGGY 378
Db 611 AAAAAAAGAGAGAGGGYGW-GDGGYGSAAAAAAGAGAGGGYGGY 669
QY 379 GT-----SGRGLGG-----QGAGAAAAAAGAGAGGGY 410
Db 670 SAAAAAAGAGAGGGYGGYCGDGGYCGSDSAAAAAAGAGAGGGY 729
QY 411 GGLGSGGT-----SGRGLGGGAGAAAAAAGAGAG--GGY-----G 452
Db 730 GSGSSAAAAAASARRACHDRAGSAAAAAAGAGAGGGYCGDGGY 789
QY 453 SGT-----SGRGLGG-----QGAGAAAAAAGAGAGGGY 486
Db 790 SDSAAAAAAGAGAGGGYGGYCGDGGYGSAAAAAAGAGAGGGY 849
QY 487 GGLGSGGT-----SGPGYGPQGTSGRGLGGGAGAAAAAAGAG 529
Db 850 GYGSDSAAAAAAGAGAGGGYGGYCGDGGYCGSDSAAAAAAGAGAG 908
QY 530 GAGGGYGLGSGT-----SGRGLGGGAGAAAAAAGAGAG--GGY 575
Db 909 GRGDGGYGSAAAAAASARRACHDRAGSAAAAAAGAGAGGGYGG 968
QY 576 GLGSGTSGRGLGGGAGAAAAAAGAGAGGGYGG----- 614
Db 969 W-----GDGGYGSAAAAAAGAGAGGGYCGDGGYCGSDSAAAA 1021
QY 615 -----LGSQGTSGRGLGGGAGAAAAAAGAGAGGGY-----G 656
Db 1022 AAAAAAGAGGGYGGYCGSDSAAAAAAGAGAGGGYCGDGGYCGSDSAAA 1081
QY 657 -----GTSGPGYGPQGTSGIRPAA 678
Db 1082 AAAAAAGAGAGGGYCGDGGYGSAAAAAAGAGAGAGAGAGAGAGAGAG 1115


```
QY 665 GPGQQTSG 672
Db 1054 DGGGGGQ 1061

RESULT 13
O53559 ID O53559 PRELIMINARY; PRT; 1489 AA.
AC O53559;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PGRS-family protein.
GN RV3514 OR MV023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL022022; CAA17751.1; -.
DR TuberculList; RV3514; -.
DR InterPro; IPR000084; PE.region.
DR DR
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE.region; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
DR Complete proteome.
KW
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBFA3A CRC64;

Query Match . 35.98; Score 1254.5; DB 16; Length 1489;
Best Local Similarity 39.5%; Pred. No. 3.7e-60;
Matches 336; Conservative 36; Mismatches 267; Indels 211; Gaps 36;

QY 16 GRGGLGGG-----AGAAAAAAGAGAGQ-----GGYGLGSGQTSGRGLGGGQAGA 65
Db 260 GHGVGDDGWLAPCGAGGAGGCGGAGAGSDGALGTTGTTG--GTGGAGGAGRGALL 317

QY 66 AAAAAAAGGA-QGGYVGLGSGQ-----GTSRGLGSGQTSGRGLGGGQAGGA 113
Db 318 LGAGGQGLGGAGGCGGTGGAGGDLVGGVGTGKGVGAGVAGLGAGGAGLFSNAG 377

QY 114 GAGGQYVGLGSGQTSRGLGGGGA-----GAAAAAAGGA-----GQGG 157
Db 378 AAGNAGVGGAGGQ--GDGGAGGAGADADQPGATGTTGTFAGGAGGAGGAGGAGGTTNG 435

QY 158 YGGLGSGQTSQPGYVGPQOOTSRRGLGGGCGAGAAAAAAGAGGQYGG-LGSGQT 216
Db 436 SGGAGGQGGAGGAGADNPITGTTGGTGGGDTGGAAGAGAGGAGGAGTGGTGMIGTTGN 495

QY 217 SRRGLGAG-----GAGAAA-----AAAAAAGGA-----GQGGYGG 248
Db 496 AGVGGAGGQGGAGGAGADADQPGATGTTGTFAGGAGGAGGAGGSSGAGTNGSGGAGG 555

QY 249 LGSQG-----TSRGGILGGGCGAGAAAAAAGAGGQYGG-LGSGQTSGR 295
Db 556 TGGGQGGAGGAGGADNPITGTTGGTGGGDTGGAAGAGGAGGAGGAGTGGTGMIGTTGNAGV 615

QY 296 GGLGGQ----CAGAAAAA-----AAAAAGAGQ-----GGYVGLGS 327

Db 616 GGAGQGGDGGAGGAGADADQPGATGTTGTFAGGAGGAGKAGSSSAGTNSGSGNGTGR 675
QY 328 Q-GTSPGGYGPQOOTSRRGLGGGCGAGAAAAAAGAGGQYGG-LGSGQTSRGG 385
Db 676 QSGTGGAGGAGADNPITGTTGGTGGGDTGGAAGAGGAGGAGTGGTGMIGTTGNAGVGG 734
QY 386 LGGQAGAAAAAAGAGGQYVGLGSGQT---SRGGILGGGCGAGAAAAAAGAGG 442
Db 735 AGSSGAGTNGSGAGGTTDQGGAGGAGGAGADNPITGTTGGTGGTGGAGAGGAGGA 794
QY 443 AGQGYGG-LGSGQTSRGGILGGQ-----GAGAAAAA-----AAAAAGGA----- 481
Db 795 AGTGTGGMIGTTGNAGVGGAGGQGGDGGAGGAGADADQPGATGTTGTFAGGAGGAGGSGG 854
QY 482 -----GQGGYGLGSGQTSRGGILGGGCGAGAAAAAAGGA-----GQGGYGLGSGQ 505
Db 855 SSCAGTNGSGAGGTCQVAGGAGTFSNSNGTGGTGGVG--GTGGDGNAGTCAG 912
QY 506 TSGRGLGGGCGAGAAAAAAGAGG--QGGYVGLGSGQTSRRGLGGGCGAGAAAAAAGGA 564
Db 913 DPGKGTGGTG--GTGGSGGAGGSGGANFNCGTGTG--GTGCKGGLNTDGLSSATSGTCG 969
QY 565 AAGAGQGGYVGLGSGQTSRGGILGGGCGAGAAAAAAGGA-----GQGGYGLGSGQ 618
Db 970 TGGTGGKGTGGAGDSDSAGTGGTGGAGGAGAGGLANTGTAGNAGTGGDGGGNGGQ 1029
QY 619 GTSRGGILGGQ-----GAGAAAAA-----AAAAAGCAGGQYGGCGLG----- 655
Db 1030 GDSG-SGLGGQPGFAGGAGGKGGAGGSGGAGGTGCGGAGGAGGAGGAGGAGISFNSGN 1088
QY 656 QTSQPGGYG 665
Db 1089 GGTGTGGVG 1098

RESULT 14
O53557 ID O53557 PRELIMINARY; PRT; 1079 AA.
AC O53557;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PGRS-family protein (Fragment).
GN RV3512 OR MV023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL022022; CAA17749.1; -.
DR TuberculList; RV3512; -.
DR InterPro; IPR002202; HMG-CoA_red.
DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
DR Complete proteome.
KW NON_TER
FT NON_TER
SQ SEQUENCE 1079 AA; 81163 MW; A79718DCB74B97D CRC64;

Query Match 35.28; Score 1228.5; DB 16; Length 1079;
```

Best Local Similarity 40.0%; Pred. No. 7.2e-59;		
Matches 318; Conservative 25; Mismatches 303; Indels 149; Gaps 29;		
Qy	14	ASRGRLGGGAGAGAAAAAAGAGAGGCGYGGGLGSGQTSRGGGLGGCGACAGAAAAAAA 73
Db	38	AGGAGGAGGTG-GTGGAAAGTTGGGQNGNGGNG--GTGKGCTGGDAGALAGSSGGAGG 94
Qy	74	AGG-----AGGGYGGGL-----SQGTSRGGGLGGQ 99
Db	95	KGGNGDGAGKAGTGSAPCTAGTGGDCRGGNGTGAAGCTGPGVTGASGCTGGSGAGGT 154
Qy	100	GA-GAAAAAAAAGGA-----GQGGYGGGL-----SQGTSRGGGLGGQ 137
Db	155	GGDGAANGGTAGAGGAGNGGKGDKAGYTSSTAGNSGGAGSGGKGGDAGAGGAGAT 214
Qy	138	-----GAGAAAAAAAAGAGAGCGYGGGLGSGQ-----GTSQPGYGPQGQTS 180
Db	215	PGANGIAGNGDGDGDAAGAVISGATCAGDGHGTTGAAGNGGTGAGGSGIDVGGG 274
Qy	181	RGGLGGCGAGAAAAAAAAGAG--QCGYGGGLGSGQ-----GTSRGGGLGGQ 225
Db	275	TGGTGGNGCAIGGAGGDGCGNSCGNGTGGKGNAGAGGAAGSGTGVANGTGGD 334
Qy	226	G--AGAAAAAAAAGAGAGCGYGGGLGSGQTSRGGGLGGQ-----AGAAAAA 272
Db	335	CGNGAGAAATAGNSGAGTGSAGNG--GTGRRGSGGAGDGTGGYGGKGGNGADGE 392
Qy	273	AAAGGAG-----QGGYGGGLGSGQTSRGGGLGGQ-----AGAAAAA 325
Db	393	VGGAGGAGGSGPNTSPGNGGQGGGSGGAG--GAAGAGAGGGGANGTAGNGGGGAGGT 451
Qy	326	GSGQTSGP---GGYGPQQTSGRGLGGCGAGAAAAAAGAGAGCGYGGGLGSGQTS 382
Db	452	GCAGAASSATNGSGGAGGTGDCGSGGAGCTGGAGGTGGAAGDGGCGGAGG--GAGG 510
Qy	383	RGGLGGQ-----AGAAAAAAAAG--AGCGYGGGLGSGQ-----GTSRGGGLGGQGA 429
Db	511	QGGAGGAGTGGNGGNTGTTGTTAGTAGAAGNGGAAGKGGAGGCGGTGGGTGGGAGGDDG 570
Qy	430	CAAAAAAAGAGAGCGYGGGLGSGQTSRGGGLGGCGAGAAAAAAGAGAGCGYGGGL 489
Db	571	AGGTGDRTVGGTVPAGSGQG--GNAGGGAGGQ--GGADGSGDGGDAGTGGNGN 626
Qy	490	G-----SQGTSRGGYGPQQTSGRGLGGCGAGAAA--AAAAAAGGAGCGYGG--GLSG 542
Db	627	GGNRNSNGTGGAGNCGGGAGGAGGAGGCGGCTGGNGAGGAGDAGDAGNGCNGTGN 686
Qy	543	GTSGRGLGGCGAGAAAAAAGAGAGCGYGGGL-----GSGQT--SGRGLGGCGAGAAA 597
Db	687	GNGNGGTAGNGGAGTGSNGNGSGGNGNAGNGSGTSCDGGAGGNGGAGT 746
Qy	598	AAAAAAGAGCGYGGGLGSGQTSRGGGLGGCGAGAAAAAAGAGAGCGYGGGLGSGQ 657
Db	747	GGTGDDGLTCTGTTGGSG--GTGGDGGNGGADNTANMTAQAGDGGNGDGGFGGGA 804
Qy	658	TSQPGYGPQQTSG 672
Db	805	GAGGGLTAGANGTG 819
RESULT 15		
Q9BIV1 PRELIMINARY; PRT; 447 AA.		
ID	Q9BIV1	AC
DT	01-JUN-2001	(TrEMBLrel. 17, Created)
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)
DT	01-OCT-2001	(TrEMBLrel. 18, Last annotation update)
DE	Major amputate spidroin 1 (Fragment).	

RESULT 15
Q9BIV1

ID	Q9HVI	PRELIMINARY;	PRT;	447 AA.
AC	Q9HVI			
DT	01-JUN-2001	(TREMblrel. 17, Created)		
DT	01-JUN-2001	(TREMblrel. 17, Last sequence update)		
DT	01-OCT-2001	(TREMblrel. 18, Last annotation update)		
DE	Major ampullate spideroin 1 (Fragment).			
GN	MASPL			
OS	Argiope aurantia.			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneolidae; Araneidae; Argiope.			
OX	NCBI_TaxID:15694;			

Search completed: December 18, 2002, 16:20:43
Job time : 48.982 secs

[illegible]

Search completed: December 18, 2002, 16:20:43
Job time : 48.982 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 ; Search time 46.1374 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-6
Perfect score: 3543
Sequence: 1 MASWTGQQMGSGMASGRG.....TSGIRAPSTSFHHHHH 691

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3543	100.0	691	22	AA82610 Spider recombinant
2	3390	95.7	681	22	AA82609 Spider recombinant
3	2431	68.6	1255	23	AA50037 N. clavipes spidro
4	2431	68.6	1880	23	AA50039 N. clavipes spidro
5	2384	67.3	989	23	AA50038 N. clavipes spidro
6	2321.5	65.5	1626	23	AAU11798 Dragline protein 1
7	2298	64.9	1617	23	AAU11794 Dragline protein 1
8	2283.5	64.5	818	23	AAU11797 Dragline protein 1
9	2254	63.6	809	23	AAU11793 Dragline protein 1
10	2151	60.7	630	23	AA50042 N. clavipes spidro

11	2151	60.7	676	23	AA50047 N. clavipes spidro
12	2138	60.3	718	19	AA53346 Nephila clavipes s
13	2138	60.3	718	21	AA53070 N. clavipes spider
14	2135	60.3	651	23	AA40097 Spider silk protei
15	2135	60.3	651	23	AAU11781 Spider natural sil
16	2129	60.1	718	12	AA14308 N.clavipes draglin
17	2076	58.6	606	16	AA99053 Spider dragline va
18	2076	58.6	606	20	AA40100 Polymer of an anal
19	2059	58.1	646	18	AA27178 Nephila clavipes s
20	1972.5	55.7	606	16	AA99055 Spider dragline va
21	1972.5	55.7	606	20	AA40102 Polymer of an anal
22	1971.5	55.6	604	16	AA99057 Spider dragline va
23	1963.5	55.4	606	20	AA40101 Polymer of an anal
24	1784.5	50.4	528	22	AA82611 Spider recombinant
25	1669.5	47.1	777	23	AA50046 N. clavipes spidro
26	1614	45.6	773	23	AA50045 N. clavipes spidro
27	1458.5	41.2	364	23	AA50043 N. clavipes spidro
28	1405.5	39.7	831	16	AA80168 pMIS1 MISP spider
29	1400	39.5	615	20	AA40099 Spider silk protei
30	1392.5	39.3	655	23	AB976672 Protein related to
31	1357	38.3	291	22	AA82608 Spider recombinant
32	1346.5	38.0	714	16	AA99059 Spider dragline va
33	1346.5	38.0	714	20	AA40103 Polymer of an anal
34	1241.5	35.0	1177	9	AA80940 SLP111 protein com
35	1241.5	35.0	1177	11	AA805307 SLP III (Silk-fibr
36	1241.5	35.0	1177	17	AA95105 Silk like protein
37	1241.5	35.0	1177	18	AA26342 SLP111 amino acid
38	1241.5	35.0	1177	21	AA78277 Repetitive protein
39	1241.5	35.0	1177	22	AA72725 SLP111 protein seq
40	1241.5	35.0	1177	22	AA63995 Silk-like protein
41	1241.5	35.0	1178	14	AA41007 SLP111-like protein
42	1241.5	35.0	1178	19	AA53518 Amino acid sequenc
43	1184.5	33.4	1023	19	AA53524 Amino acid sequenc
44	1184.5	33.4	1059	14	AA41013 SLP4 multimeric pr
45	1184.5	33.4	1059	18	AA26348 SLP4 synthetic pro

ALIGNMENTS

RESULT 1
AA82610
ID AA82610 standard; Protein; 691 AA.
XX
AC AA82610;
XX
XX 02-OCT-2001 (first entry)
XX
DE Spider recombinant silk protein pET((SP1)4/(SP2)1)4.
DE
KW Spider; orb-weaver; silk protein; pET((SP1)4/(SP2)1)4;
KW structural protein; purification; fibre; spinning.
XX
OS Nephila clavipes.
XX
XX WO200153333-A1.
PN
XX 26-JUL-2001.
PD
XX 01-NOV-2000; 2000WO-US30086.
XX
XX 20-JAN-2000; 2000US-0490291.
PR
XX (MELL/) MELLO C M.
PA (ARCI/) ARCIDIAcono S.
PA (BUTL/) BUTLER M M.
PA (USSA) US SEC OF ARMY.
XX
XX Mello CM, Arcidiacono S, Butler MM;
PI WPI; 2001-483136/52.
XX
DR N-PSDB; AA826303.
DR
XX


```
Db 421 GAGAAAAAGGAGGGGGLGSGAGGGGAGAAAAAGAGGGGGLGSGAGGGG 480
Qy 295 -----SRRGGLGGGAGAAAA----- 310
Db 481 GAGAAAAAGGAGGGGGLGSGAGGGGAGGGGGLGSGAGGGGGLGSGAGGGG 540
Qy 311 -----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 340
Db 541 GGLGGGAGGGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 600
Qy 341 GQQTSGRGLGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 398
Db 601 GSGAGRGLGGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 656
Qy 399 AAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 429
Db 657 AAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 716
Qy 430 GAGAAAA-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 463
Db 717 GAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 775
Qy 464 LGGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 522
Db 776 LGGGGAG-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 830
Qy 523 AA-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 546
Db 831 AAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 890
Qy 547 -----SRRGGLGGGAGAAAA----- 562
Db 891 AGGGGGLGGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 950
Qy 563 -----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 600
Db 951 AGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLG 1009
Qy 601 -----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 638
Db 1010 GGCGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 1068
Qy 639 -----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 669
Db 1069 GAGGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 1108

RESULT 5
ID AAM50038
AC AAM50038 standard; Protein; 989 AA.
XX AAM50038;
XX AAM50038;
DT 18-SEP-2002 (first entry)
DE N. clavipes spideroin synthetic homologue SO1SM12 protein.
KW Spideroin; spider; silk; fibre; film; membrane; wound; filter; SO1SM12.
OS Synthetic.
PN DE10113781-Al.
XX DE10113781-Al.
PD 13-DEC-2001.
XX 13-DEC-2001.
XX 21-MAR-2001; 2001DE-1013781.
XX 09-JUN-2000; 2000DE-1028212.
XX 24-OCT-2000; 2000DE-1053478.
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX Scheller J, Conrad U, Grosse F, Guehrs K;
PI
```

```
XX WPI; 2002-123561/17.
DR N-PSDB; ABL61039.
XX
PT New DNA encoding synthetic spider silk protein, useful e.g. for closing
PT wounds, comprises modules that encode repeating units of spideroin
PT proteins
XX
PS Claim 22; Page 34-38; 88pp; German.
XX
CC This invention describes a novel DNA sequence, encoding a synthetic
CC spider silk protein, comprising modules, each comprising a group of
CC sequentially arranged oligonucleotides, each oligonucleotide encoding
CC a repeating unit of a spideroin protein. The synthetic protein has at
CC least 84% homology with the Nephila clavipes spideroin protein and is used
CC to produce synthetic fibres, films and/or membranes, particularly: (i)
CC for medical use, especially to close wounds and/or to support or cover
CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
CC (iii) as filters. The synthetic proteins are very similar to native
CC spider silk proteins; can be prepared on a large scale and can be spun to
CC fibres with excellent mechanical properties (strength and elasticity).
CC Also they retain water solubility after long-term boiling in aqueous
CC solutions and since they are also soluble in organic solvents but
CC precipitated at high salt concentration, they are easily extracted and
CC purified. The modular construction of the invention facilitates
CC incorporation of additional peptide-encoding sequences, e.g. to simplify
CC purification or modulate solubility. This sequence represents the
CC synthetic N. clavipes spideroin-1 homologue SO1SM12 described in the
CC invention.
XX
SQ Sequence 989 AA;
```

```
Query Match 67.3%; Score 2384; DB 23; Length 989;
Best Local Similarity 57.8%; Pred. No. 5.8e-163;
Matches 565; Conservative 18; Mismatches 38; Indels 356; Gaps 35;

Qy 6 GGQMGRRSMASRRGGLGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 65
Db 9 GGQAGGG-----GYGGLGGGAGGGGAGAAAAAGGAGGGGGLGSGAGGGG 60
Qy 66 GAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 108
Db 61 G-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 115
Qy 109 -----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 159
Db 116 AGGGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 170
Qy 160 YGGLGSGGTSPPGYP-GQQTSGRGLGGGAGAAAAAGGAGGGGGLGSGAGGGG 198
Db 171 YGGLGSGG-AGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 229
Qy 199 AAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 218
Db 230 AAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 289
Qy 219 --SRRGGLGGGAGAAAA-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 256
Db 290 QGAGRGLGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 348
Qy 257 SRRGGLGGGAGAAAA-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 290
Db 349 AGRGLGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 408
Qy 291 SGGTSGRGLGGGAGAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 345
Db 409 SGG-AGRGLGGGAG-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 462
Qy 346 ---GRGGLGGGAG---GAAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 387
Db 463 GGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGGGLGSGAGGGG 522
Qy 388 LGGGGAGAAAA-----AAAAAGGAGGGGGLGSGAGGGGGLGSGAGGGG 427
```


KW Dragline protein 1; DP-1B 8mer; His tag.
XX Nephila clavipes.
OS Synthetic.
XX WO200190389-A2.
XX PN 29-NOV-2001.
XX PD 24-MAY-2001; 2001WO-US16937.
XX PF 25-MAY-2000; 2000US-206968P.
XX PR (DUPO) DU PONT DE NEMOURS & CO E I.
XX PA Yang JG;
XX PI WPI; 2002-106209/14.
XX DR Producing silk-like proteins in a green plant, useful in fabrics and
XX PT material construction, comprises providing a silk-like protein
XX PT expression cassette to a green plant -
XX PS Example 1; Page 83-86; 93pp; English.
XX CC The invention relates to a method of producing silk-like proteins (SLP)
CC in a green plant by providing a green plant containing an SLP expression
CC cassette. The green plant contains the expression cassette P-SLP-T where
CC P is a promoter for driving the expression of an SLP transgene and T is a
CC 5' terminator. The method is useful for producing silks and silk-like
CC proteins in green plants. The silks and silk-like proteins may be used in
CC fabrics or in material construction, such as rope, surgical sutures,
CC flexible tie downs for certain electrical components, or as a
CC biomaterial for implantation. The method allows for more cost effective
CC production of silk not obtained from natural or microbial sources.
CC The present sequence is the silk protein Dragline protein (DP) 1
CC synthetic variant DP1B/his tag 8mer (i.e. 8 copies of DP1B with a His tag
CC at the C-terminus) which is used as an SLP construct in the method of the
CC invention.
XX SQ Sequence 818 AA:

Query Match 64.5%; Score 2283.5; DB 23; Length 818;
Best Local Similarity 59.6%; Pred. No. 7.5e-156;
Matches 544; Conservative 11; Mismatches 41; Indels 317; Gaps 36;

QY 1 MASMTGGQGMGRGSMAS---GRGLGGGQAGAAAA-----AAAA 36
DB 1 MARSQAGGGYGGGLGSGQAGRGGLGGQAGAAAAAAGAGAGGQGGGLGSGQAGGAGAAAA 60
QY 37 AAGGAGGGYGGGLGSGQTSRGGGLGGQAGAAAAAAGAGAGGQGGYGGGLGSGQGT---- 92
DB 61 AAGGAGGGYGGGLGSGQ----AGRGQAG----AAAAAAGAGGQGGYGGGLGSGQAGGG 112
QY 93 -----SRGGLGGQAGAAAA-----AAAAAAGAGGQGGYGG 123
DB 113 YGGLSGQAGRGGLGGQAGAAAAAAGAGAGGQGGGLGSGQAGGAGAAAAAAGAGGQGGY 172
QY 124 GLGSGQTSRGGGLGGQAGAAAAAAGAGAGGQGGYGGGLGSGQTSRGGYGP-GQOTS 182
DB 173 GLGSGQ----AGRGQAG----AAAAAAGAGGQGGYGGGLGSGQ-AGQGGYGGGLGSGQAG 223
QY 183 RGLGGQAGAAAA-----AAAAAAGAGGQGGYGGGLGSGQTSGR 221
DB 224 RGLGGQAGAAAAAAGAGGQGGGLGSGQAGGAGAAAAAAGAGGQGGYGGGLGSGQ---- 279
QY 222 GGLGGQAGAAAAAAGAGGQGGYGGGLGSGQGT-----SRGGLGGQAG 269
DB 280 AGRGGQAG----AAAAAAGAGGQGGYGGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAG 335
QY 270 AAA-----AAAAAAGAGGQGGYGGGLGSGQTSRGGGLGGQAGAA 308
DB 336 AAAAAGAGGQGGGLGSGQAGGAGAAAAAAGAGGQGGYGGGLGSGQ----AGRGQAG--- 389

QY 309 AAAAAAAGAGGQGGYGGGLGSGQTSRGGYGP-GQOTSRRGGLGGQAGAAAA----- 360
DB 390 --AAAAAAGAGGQGGYGGGLGSGQ-AGQGGYGGGLGSGQAGRGGLGGQAGAAAAAAGGAG 446
QY 361 -----AAAAAAGAGGQGGYGGGLGSGQTSRGGGLGGQAGAAAAAAGG 406
DB 447 QGGLGSGQAGGAGAAAAAAGAGGQGGYGGGLGSGQ----AGRGQAG----AAAAAGG 498
QY 407 AGQGGYGGGLGSGQGT-----SRGGLGGQAGAAAA----- 436
DB 499 AGQGGYGGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAGAAAAAAGAGGQGGGLGSGQAG 558
QY 437 ---AAAAAAGAGGQGGYGGGLGSGQTSRGGGLGGQAGAAAAAAGAGGQGGYGGGLG 493
DB 559 GAGAAAAAAGAGGQGGYGGGLGSGQ----AGRGQAG----AAAAAAGAGGQGGYGGGLG 610
QY 494 QGTSRGGYGP-GQOTSRRGGLGGQAGAAAA-----AAAAAG 531
DB 611 QG-AGQGGYGGGLGSGQAGRGGLGGQAGAAAAAAGAGGQGGGLGSGQAGGAGAAAAAG 669
QY 532 GAGQGGYGGGLGSGQTSRGGGLGGQAGAAAAAAGAGGQGGYGGGLGSGQGT----- 584
DB 670 GAGQGGYGGGLGSGQ----AGRGQAG----AAAAAAGAGGQGGYGGGLGSGQAGGQGGY 721
QY 585 -----SRGGLGGQAGAAAA-----AAAAAAGAGGQGGYGG 618
DB 722 LGSQAGRGGLGGQAGAAAAAAGAGGQGGGLGSGQAGGAGAAAAAAGAGGQGGYGG 781
QY 619 SGTSGRGGGLGGQAGAAAAAAGAGGQGGYGGGLGSGQTSRGGYGP-GQOTSIRIR 678
DB 782 SQG----AGRGQAG----AAAAAAGAGGQGGYGGGLGSG----- 812
QY 679 APSTSEFHHHHH 691
DB 813 -----HHHHH 818

RESULT 9
AAU11793
ID AAU11793 standard; protein; 809 AA.
XX AC AAU11793;
XX DT 26-MAR-2002 (first entry)
XX DE Dragline protein 1 analogue DP-1B 8mer.
XX KW Orb-weaving spider; silk-like protein; SLP; transgenic plant;
XX KW promoter; 5' terminator; fabric production; material construction; rope;
XX KW surgical suture; flexible tie down; electrical component; implantation;
XX KW Dragline protein 1; DP-1B 8mer.
OS Nephila clavipes.
OS Synthetic.
XX PN WO200190389-A2.
XX PD 29-NOV-2001.
XX PF 24-MAY-2001; 2001WO-US16937.
XX PR 25-MAY-2000; 2000US-206968P.
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Yang JG;
XX DR WPI; 2002-106209/14.
XX PT Producing silk-like proteins in a green plant, useful in fabrics and
XX PT material construction, comprises providing a silk-like protein
XX PT expression cassette to a green plant -

XX Disclosure; Page 74-76; 93pp; English.

CC The invention relates to a method of producing silk-like proteins (SLP) in a green plant by providing a green plant containing an SLP expression cassette. The green plant contains the expression cassette P-SLP-T where P is a promoter for driving the expression of an SLP transgene and T is a 5' terminator. The method is useful for producing silks and silk-like proteins in green plants. The silks and silk-like proteins may be used in fabrics or in material construction, such as rope, surgical sutures, CC flexible tie downs for certain electrical components, or as a CC biomaterial for implantation. The method allows for more cost effective CC production of silk not obtained from natural or microbial sources. CC The present sequence is the silk protein dragline protein (DP) 1 CC synthetic variant DP1B 8mer (i.e. 8 copies of DP1B) which is used as CC an SLP construct in the method of the invention.

XX SQ Sequence 809 AA;

Query Match 63.6%; Score 2254; DB 23; Length 809;
 Best Local Similarity 61.6%; Pred. No. 9.5e-154;
 Matches 533; Conservative 11; Mismatches 35; Indels 286; Gaps 34;

QY 11 GRGSMASRGGLGCGAGAAA-----AAAAAGAGGCGGCGGL 49
 DB 12 GLGSGAGRGGLGCGAGAAAAGAGAGCGGLGSGAGAGAGAGCGGCGGL 71

QY 50 GSQTSRGLGCGAGAGAAAAGAGAGCGGCGGCGGLGSGT-----SGRG 97
 DB 72 GSQC-----AGRGCGAG-----AAAAAGAGCGGCGGCGGCGAGCGG 123

QY 98 LGGCGAGAAA-----AAAAAGAGCGGCGGCGGLGSGT-----AGR 179
 DB 124 LGGCGAGAAAAGAGAGCGGCGGCGAGAGAGAGAGCGGCGGCGGCGG 195

QY 137 GGCGAGAGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 234
 DB 180 GGCGAG-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 234

QY 196 A-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 286
 DB 235 AAAAG 272

QY 235 AAAAG 346
 DB 273 -----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 321

DB 347 LGSOGAG 398

QY 322 GYGGLGSGTSGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 DB 399 GYGGLGSGTSGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457

QY 361 -AAAAAG 419
 DB 458 GAAAG 509

QY 420 T-----SCRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 446
 DB 510 AGGCGGGLGCGAG 569

QY 447 QGCGGGLGSGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 505
 DB 570 QGCGGGLGSGT-----AGRGCGAG-----AAAAAGAGAGAGAGAGAGAG 620

QY 506 QGCGGGLGCGAG 544
 DB 621 SQAGRGGLGCGAG 680

QY 545 GTSRGLGCGAG 592

DB 681 G-----AGRGCGAG-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 732

QY 593 QGAG 631

DB 733 QGAG 788

QY 632 GAG 656

DB 789 GAG-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809

RESULT 10
 AAM50042
 ID AAM50042 standard; Protein; 630 AA.
 AC AAM50042;
 XX 18-SEP-2002 (first entry)
 DT N. clavipes spidroin synthetic homologue S01 protein.
 DE Spidroin; spider; silk; fibre; film; membrane; wound; filter; S01.
 KW Synthetic.
 OS DEL0113781-AL.
 PN 13-DEC-2001.
 PD 21-MAR-2001; 2001DE-1013781.
 PF 09-JUN-2000; 2000DE-1028212.
 PR 24-OCT-2000; 2000DE-1053478.
 XX (IPK-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
 PA Scheller J, Conrad U, Grosse F, Guehrs K;
 PI WPI; 2002-123561/17.
 DR N-PSDB; ABL61043.
 XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
 PT wounds, comprises modules that encode repeating units of spidroin
 PT proteins
 XX Claim 22; Page 48-50; 88pp; German.

CC This invention describes a novel DNA sequence, encoding a synthetic
 CC spider silk protein, comprising modules, each comprising a group of
 CC sequentially arranged oligonucleotides, each oligonucleotide encoding
 CC a repeating unit of a spidroin protein. The synthetic protein has at
 CC least 84% homology with the Nephila clavipes spidroin protein and is used
 CC to produce synthetic fibres, films and/or membranes, particularly: (i)
 CC for medical use, especially to close wounds and/or to support or cover
 CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
 CC (iii) as filters. The synthetic proteins are very similar to native
 CC spider silk proteins; can be prepared on a large scale and can be spun to
 CC fibres with excellent mechanical properties (strength and elasticity).
 CC Also they retain water solubility after long-term boiling in aqueous
 CC solutions and since they are also soluble in organic solvents but
 CC precipitated at high salt concentration, they are easily extracted and
 CC purified. The modular construction of the invention facilitates
 CC incorporation of additional peptide-encoding sequences, e.g. to simplify
 CC purification or modulate solubility. This sequence represents the
 CC synthetic N. clavipes spidroin-1 homologue S01 described in the
 CC invention.

XX SQ Sequence 630 AA;

Query Match 60.7%; Score 2151; DB 23; Length 630;
 Best Local Similarity 58.6%; Pred. No. 1.8e-146;
 Matches 497; Conservative 9; Mismatches 41; Indels 178; Gaps 32;

QY	6	GGQOMGRGSMASGRGGLGGGACAGAAAAAAGGAGCGGCGGGLGSGQTSGRGGLGGQGA	65
Db	9	GGGAGGQGG-----GYGLGGQGGAGAGAAAAAGGAGCGGCGGGLGSGQ-----AGRGQGA	60
QY	66	GAIAAAAAAGAGCGGCGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGCGGCGGGLG	125
Db	61	G-----AAAAAGAGCGGCGGGLGSGQ-AGRGGLGGQAG-----AAAAAGCAGCGGCGGGL	111
QY	126	GSQGTSGRGLGGGAGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGYGPQGTSGRG	185
Db	112	GSQG-----AGRGQGGAG-----AAAAAGGAGCGGCGGGLGSGQ-----AGRG	150
QY	186	LGSGGAGAAAAA-----AGGAGCGGCGGGLGSGQTSRGLGGGAGAAAAA	236
Db	151	LGQGGAGAAAAAGGAGCGGCGGGLGGGAGCGGCGGGLGSGQ-AGRGGLGSGAG-----AA	205
QY	237	AAAGGAGCGGCGGGLGSGQTSRGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTS	296
Db	206	AAAGGAGQ-----GGLGGGAGCGGAGAAAAAGGAGCGGCGGGLGSGQ-----	248
QY	297	RGGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGYGP-QQGTSGRGLGGQGA	355
Db	249	AGRGQGGAG-----AAAAAGGAGCGGCGGGLGSGQ-AGQGGYGGGLGSGQAGRGGLGGQGA	302
QY	356	GAIAA-----AAAAAGCAGCGGCGGGLGSGQTSRGLGGGAG	395
Db	303	GAIAAAGGAGCGGGLGGGAGAGAGAAAAAGGAGCGGCGGGLGSGQ-AGRGGLGGQAG	360
QY	396	AAAAAAGGAGCGGCGGGLGSGQTSRGLGGGAGAAAAAAGGAGCGGCGGGLG	455
Db	361	-----AAAAAGGAGCGGCGGGLGSGQ-----AGRGQGGAG-----AAAAAGCAGCGGCGGGLG	409
QY	456	QGTSGRGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGYGPQGTSGRGLG	515
Db	410	QG-AGRGGLGGQAG-----AAAAAGGAGCGGCGGGLGSGQ-----AGRG	448
QY	516	GCGAGAAAAAAGGAGCGGCGGGLGSGQTSRGLGGGAGAAAAA	567
Db	449	GQAG-----AAAAAGGAGCGGCGGGLGSGQ-----AGRGQGGAGAAAAAGGAGCGGCGGGL	500
QY	568	AGGAGCGGCGGGLGSGQTSRGLGGGAGAAAA-----AAAAA	606
Db	501	GCGAGCGGCGGGLGSGQ-AGRGGLGGGAGAAAAAGGAGCGGGLGGGAGCGGAGAAAAA	559
QY	607	GAGCGGCGGGLGSGQTSRGLGGGAGAAAAAAGGAGCGGCGGGLGSGQ-----SGPG	664
Db	560	GAGCGGCGGGLGSGQ-----AGRGQGGAG-----AAAAAGGAGCGGCGGGLGSGQAGRGGLG	611
QY	665	GYPGP	669
Db	612	GQAG	616
RESULT 11			
AAW50047			
ID	AAW50047 standard; Protein; 676 AA.		
XX	AAW50047;		
AC			
XX			
DT	18-SEP-2002 (first entry)		
XX			
DE	N. clavipes spidroin synthetic homologue SOL protein #2.		
XX			
KW	Spidroin; spider; silk; fibre; film; membrane; wound; filter; SOL.		
OS	Synthetic.		
XX			
FH	Key		
FT	Peptide		
FT	1..28		
FT	/label= LeB4_signal_peptide		
FT	29..659		
FT	/note= "Synthetic spidroin homologue SOL"		
FT	660..676		

FT	Domain	/note= "c-myc-tag" 673..676
FT		/note= "ER retention signal"
XX	DE10113781-A1.	
XX	13-DEC-2001.	
PD		
XX	21-MAR-2001; 2001DE-1013781.	
PF		
XX	09-JUN-2000; 2000DE-1028212.	
PR		
PR	24-OCT-2000; 2000DE-1053478.	
XX		
PA	(IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.	
XX		
PI	Scheller J, Conrad U, Grosse F, Guehrs K;	
XX	WPI; 2002-123561/17.	
DR		
XX		
PT	New DNA encoding synthetic spider silk protein, useful e.g. for closing	
PT	wounds, comprises modules that encode repeating units of spidroin	
PT	proteins	
XX		
PS	Example 1; Fig 10A; 88pp; German.	
XX		
CC	This invention describes a novel DNA sequence, encoding a synthetic	
CC	spider silk protein, comprising modules, each comprising a group of	
CC	sequentially arranged oligonucleotides, each oligonucleotide encoding	
CC	a repeating unit of a spidroin protein. The synthetic protein has at	
CC	least 84% homology with the Nephila clavipes spidroin protein and is used	
CC	to produce synthetic fibres, films and/or membranes, particularly: (i)	
CC	for medical use, especially to close wounds and/or to support or cover	
CC	artificial organs; (ii) as adhesion surfaces for culturing cells; and	
CC	(iii) as filters. The synthetic proteins are very similar to native	
CC	spider silk proteins; can be prepared on a large scale and can be spun to	
CC	fibres with excellent mechanical properties (strength and elasticity).	
CC	Also they retain water solubility after long-term boiling in aqueous	
CC	solutions and since they are also soluble in organic solvents but	
CC	precipitated at high salt concentration, they are easily extracted and	
CC	purified. The modular construction of the invention facilitates	
CC	incorporation of additional peptide-encoding sequences, e.g. to simplify	
CC	purification or modulate solubility. This sequence represents a construct	
CC	composed of the LeB4 signal peptide, N. clavipes spidroin-1 synthetic	
CC	homologue SOL, a c-Myc-tag and an endoplasmic reticulum (ER)-retention	
CC	signal described in the invention.	
XX		
SQ	Sequence	676 AA;

Query Match	60.7%	Score 2151;	DB 23;	Length 676;
Best Local Similarity	68.6%	Pred. No. 1.9e-146;		
Matches 497;	Conservative 9;	Mismatches 41;	Indels 178;	Gaps 32;
QY	6	GGQOMGRGSMASGRGGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSRGGGLGGQGA	65	
Db	37	GGGAGCGG-----GYGGLGGGAGCGAGAAAAAGGAGCGGCGGGLGSGQ-----AGRGQGA	88	
QY	66	GAIAAAAAAGAGCGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGGAGCGGCGGGL	125	
Db	89	G-----AAAAAGGAGCGGCGGGLGSGQ-AGRGGLGGQAG-----AAAAAGGAGCGGCGGGL	139	
QY	126	GSQGTSGRGLGGGAGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGGYPGQGTSGRG	185	
Db	140	GSQG-----AGRGGQAG-----AAAAAGGAGCGGCGGGLGSGQ-----AGRG	178	
QY	186	LGCGGAGAAAAA-----AGGAGCGGCGGGLGSGQTSRGGGLGGGAGAAAAA	236	
Db	179	LGCGGAGAAAAAAGGAGCGGCGGGLGGGAGCGGCGGGLGSGQ-AGRGGLGGQAG-----AA	233	
QY	237	AAAGGAGCGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSG	296	
Db	234	AAAGGAGQ-----GGLGGGAGCGAGAAAAAAGGAGCGGCGGGLGSGQ-----	276	
QY	297	RGGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGGYGP-QGQTSRGGGLGGQGA	355	


```

      ||| | | :|| ||||| ||||| | ||||| ||||| ||||| |||||
Db 476 GYGGLGSGAGR---GGGAGAAAAAVGAGQEGIRGQAGQGQGGYGLGSG-SGRGL 531
QY 553 GGAGAGAAA-----AAAAAGGAGQGGYGLGSGQTSGRGLG 592
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 GGAGAGAAAAAGGAGQGGGLGGAGQAGAGAAAAAGGVQGGYGLGSG-----AGRG 587
QY 593 QGAGAAAAAGGAGQGGYGLGSGQTSGRGLGQAGAGAAAAAGGAGQGGY 652
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 QGAG-----AAAAAGGAGQGGYGLGGQGV-CRGLGGQAG-----AAAAGGAGQGGY 636
QY 653 GGLGS 657
      ||:|
Db 637 GGVS 641
```

Search completed: December 18, 2002, 16:17:19
Job time : 53.1374 secs

; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-19

Query Match 60.3%; Score 2135; DB 4; Length 651;
Best Local Similarity 68.3%; Pred. No. 5.7e-147;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
QY 25 QGAGAAAAAAGAGAGAGGGTGGTGGT--SGRGGGAGAG-AAAAAAGAGAG 81
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 QGAG-----AAAAAGAGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 GYGGLGGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57 GYGGLGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 142 GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 G-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 202 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 253 SGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 SGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 292 QGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 QG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 351 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 393 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 AG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 453 LGSQGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 LGNQG-AGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 500 GYGYP-GQGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 GYGGLGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 553 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 593 QGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 QGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 653 GGLGS 657
   |||||
Db 637 GGVGS 641
   |||||
```

RESULT 5

US-09-247-806-1
; Sequence 1, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude

; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-247-806-1

Query Match 60.3%; Score 2135; DB 4; Length 651;
Best Local Similarity 68.3%; Pred. No. 5.7e-147;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
QY 25 QGAGAAAAAAGAGAGAGGGTGGTGGT--SGRGGGAGAG-AAAAAAGAGAG 81
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 QGAG-----AAAAAGAGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 GYGGLGGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57 GYGGLGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 142 GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 G-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 202 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 253 SGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 SGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 292 QGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 QG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 351 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 393 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 AG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 453 LGSQGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 LGNQG-AGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 500 GYGYP-GQGTSGRGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 GYGGLGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 553 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 593 QGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 QGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 653 GGLGS 657
   |||||
Db 637 GGVGS 641
   |||||
```


RESULT 14
US-08-556-978B-61
; Sequence 61, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-61

Query Match 38.0%; Score 1346.5; DB 4; Length 714;
Best Local Similarity 50.0%; Pred. No. 3.4e-90;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

QY 2 ASMTGCGQMGSGMASGRGLGGCGAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 10 AAAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAASAAAAAAGPGYGPQGGPGGSA 69
QY 57 RGLGGCGAGAAAAAAAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 70 PQQGPSPGSAASAAAAAAGPGYGPQGGPGGYG-PQGGPGGYGPGQGGPGGSA 128
QY 112 AAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGGGYGLGSGQTS 170
Db 129 AAAAGPGQGGPGGYG-PQGGPGGYGPGQGGPGGSAASAAAAAAGPGYG----PGQG 183
QY 171 PGYGPQQTSGRGLGGGAGAAAAAAAG----GAGGGYGLGSGQTSRGLG- 225
Db 184 PGYGPQQTSGRGLGGGAGAAAAAAGPGYGPQGGPGGYG-PQGGPGGYG 235
QY 226 GQAGAAAAAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGAGG 284
Db 236 GQGGPGGSAASAAAAAAGPGYGPQGGPGGYG-PQGGPGGYGPGQGGPGGSA 294
QY 285 YGGLGSGQTSRGLGGGAGAAAAAAGAGGGYGLGSGQTSRGLGSGQTS 344
Db 295 YGYP-GQGGPGGYG-PQGGPGGSAASAAAAAAGPGYG----PGQGPGGYGPGQGG 348

QY 345 SGRGLGGCGAGAAAAAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 403
Db 349 PGYGPQGGPGGPGSAASAAA-GPQGGPGGYG-PQGGPGGYGPGQGGPGGSA 406
QY 404 AGGAGGGYGLGSGQTSRGLGGGAGAAAAAAGAGGGYGLGSGQTSRGLG- 452
Db 407 AAAAGPGGYG-PQGGPGGYG-PQGGPGGSAASAAAAAAGPGYGP-GQGGPGGYG 463
QY 463 ---GLGGCGAG-----AAAAAAGAGGGYG-----GLGSGQTS----- 497
Db 464 GQGGPGGYGPGQGGPGGSAASAAAAAAGPGQGGPGGYGPGQGGPGGSA 523
QY 498 -----GPGGYGPGQGGPGGSAASAAAAAAGAGGGYGLGSGQTSRGLG- 550
Db 524 AAAAAGPGYGPQGGPGGYGPGQGGPGGSAASAAAAAAGPGYGP-GQGGPGGYG 582
QY 551 ---GLGGCGAG-----AAAAAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 602
Db 583 GQGGPGGYGPGQGGPGGSAASAAAAAAGPGQGGPGGYG-PQGGPGGYGPGQGGPGGSA 641
QY 603 AAAAGAGGGYG-PQGGPGGYG-PQGGPGGSAASAAAAAAGAGGGYGLGSGQTSRGLG- 657
Db 642 AAAAAGPGGYGPGQGGPGGYGPGQGGPGGSAASAAAAAAGPGYGPQGGPGGYG- 700
QY 658 QGTSGPGGYGPGQ 671
Db 701 PGQGGPGGYGPGQ 714
RESULT 15
US-09-247-806-10
; Sequence 10, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-247-806-10

Query Match 38.0%; Score 1346.5; DB 4; Length 714;
Best Local Similarity 50.0%; Pred. No. 3.4e-90;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

QY 2 ASMTGCGQMGSGMASGRGLGGCGAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 56
Db 10 AAAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAASAAAAAAGPGYGPQGGPGG 69
QY 57 RGLGGCGAGAAAAAAAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 70 PQQGPSPGSAASAAAAAAGPGYGPQGGPGGYG-PQGGPGGYGPGQGGPGGSA 128
QY 112 AAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGGGYGLGSGQTS 170
Db 129 AAAAGPGQGGPGGYG-PQGGPGGYGPGQGGPGGSAASAAAAAAGPGYG----PGQG 183
QY 171 PGYGPQQTSGRGLGGGAGAAAAAAAG----GAGGGYGLGSGQTSRGLG- 225

Db 184 PGYGPQQ-----GPSGPGSAAAAAAGPGYGPQQGPGGYG-PQQGPGGYGP 235
QY 226 GQAGAAAAAAGAGGGYGLGSGTSGRGLG-CQGAGAAAAAAGAGGAGQG 284
Db 236 GQGGPSGPGSAAAAAAGPGQGGPGGYG-PQQGPGGYGPQQGPGSAAAAAAGPG 294
QY 285 YGGLGSGTSGRGLGCGAGAAAAAAGAGGAGGCGGLGSGTSGPGYGPQQQT 344
Db 295 GYGP-GQQGPGGYGP-CQQGPGPGSAAAAAAGPGGYG----PQQGPGGYGPQQG 348
QY 345 SGRGLGCGAGAAAAAAGAGGAGGCGYGLGSGTSGRGLG-CQGAGAAAAAAG 403
Db 349 PGYGPQQGPGSAAAAAAG-GPGQGGPGGYG-PQQGPGGYGPQQGPGSAAAA 406
QY 404 AGNAGGCGYGLGSGTSGRGLGCGAGAAAAAAGAGGAGGCGYGLGSGTSGRG- 462
Db 407 AAAAGPGGYGP-GQQGPGGYGP-GQQGPGSAAAAAAGPGGYGP-GQQGPGGYGP 463
QY 463 ---GLGGGAG-----AAAAAAGAGGAGGCGYG-----GLGSGTSG----- 497
Db 464 QQQPGGYGPQQGPGSAAAAAAGPGQGGPGGYGPQQGPGGYGPQQGPGSAA 523
QY 498 -----GPGYGPQQTSGRGLGCGAGAAAAAAGAGGAGGCGYGLGSGTSGRG- 550
Db 524 AAAAAAGPGYGPQQGPGGYGPQQGPGSAAAAAAGPGGYGP-GQQGPGGYGP 582
QY 551 ---GLGGGAG----AAAAAAGAGGAGGCGYGLGSGTSGRGLG-CQGAGAAAA 602
Db 583 GQQGPGGYGPQQGPGSAAAAAAGPGQGGPGGYG-PQQGPGGYGPQQGPGSA 641
QY 603 AAAAGGAGGCGYG-GLGSGTSGRGLGCGAGAAAAAAG-----GAGGCGYGLG 657
Db 642 AAAAAAGPGGYGPQQGPGGYGPQQGPGSAAAAAAGPGGYGPQQGPGGYG- 700
QY 658 QGTSGPGGYGPQQ 671
Db 701 PGQGGPGGYGPQQ 714

This Page Blank (uspto)


```
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
; US-09-861-597-10

Query Match      38.0%; Score 1346.5; DB 10; Length 714;
Best Local Similarity 50.0%; Pred. No. 1.3e-73;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

Qy  2 ASMTGGGQMGKSGMASRGLGCGAG-----AAAAAAGAGAGGAGGCGY-GLGSQGTSG 56
Db  10 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69
Qy  57 RGLGCGCGAGAAAAAAGAG-----GAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 111
Db  70 PQQGGPSGPGSAAAAAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 128
Qy  112 AAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 170
Db  129 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 183
Qy  171 PGYGPQGTSGRGLGCGAGAAAAAAGAG-----GAGCGGCGGCGCGCGCGCGCGCGCG 225
Db  184 PGYGPQGTSGRGLGCGAGAAAAAAGAG-----GAGCGGCGGCGCGCGCGCGCGCGCG 235
Qy  226 GCGAGAAAAAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 284
Db  236 GCGAGAAAAAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 294
Qy  285 GYGLGSGGTSGRGLGCGAGAAAAAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 344
Db  295 GYGP-GGCGPGGYG-GGCGPSGCGSAAAAAAGAGAGCGGCG-----PGCGGCGCGCG 348
Qy  345 SGRGLGCGAGAAAAAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 403
Db  349 PGYGPQGTSGRGLGCGAGAAAAAAGAG-----GAGCGGCGGCGCGCGCGCGCGCGCG 406
Qy  404 AGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 462
Db  407 AAGAGPGGYG-GGCGPGGYG-GGCGPSGCGSAAAAAAGAGAGCGGCG-----PGCGG 463
Qy  463 ---GLGCGAG-----AAAAAAGAGAGCGGCG-----GLGSGGCGS----- 497
Db  464 GCGGPGGYGPGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Qy  498 -----GPGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
Db  524 AAAAAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
Qy  551 ---GLGCGAG-----AAAAAAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 602
Db  583 GCGGPGGYGPGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
Qy  603 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
Db  642 AAAAAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 700
Qy  658 GGTSGPGGYGPGCG 671
Db  701 PGCGGPGGYGPGCG 714

RESULT 6
US-09-861-597-2
```

```
; Sequence 2, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Nephila clavipes
; US-09-861-597-2

Query Match      31.0%; Score 1098.5; DB 10; Length 529;
Best Local Similarity 44.4%; Pred. No. 4.9e-59;
Matches 309; Conservative 25; Mismatches 159; Indels 203; Gaps 33;

Qy  6 GCGQMGKSGMASRGLGCGAGAAAAAAGAGAGAGGAGGCGGCGCGCGCGCGCGCGCG 65
Db  5 GCGQMGKSGMASRGLGCGAGAAAAAAGAGAGAGGAGGCGGCGCGCGCGCGCGCGCG 52
Qy  66 CAAAAAAGAGAGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 123
Db  53 -----PGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 92
Qy  124 -----GLGSGGTSGRGLGCGAGAAAAAAGAGAGGCGGCGCGCGCGCGCGCGCG 176
Db  93 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 141
Qy  177 GCGTSGRGLG-----GCGAGAAAAAAGAGAGAGGAGGCGGCGCGCGCGCGCGCG 280
Db  142 GCG-----GPGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 198
Qy  226 GCGAGAAAAAAGAGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 254
Db  199 GCGSAAAAAAGAGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 254
Qy  281 AGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 340
Db  255 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 298
Qy  341 GCGTSGRGLGCGAGAAAAAAGAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 400
Db  299 GCG-----GPGGAGSAAAAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
Qy  401 AAAAAAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
Db  327 ---PGCGGCGGCG-----PGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
Qy  460 GRGLGCGAGAAAAAAGAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519
Db  376 GCG-----SASAAAAAAGAGGCGGCG-----PGCGGCGGCGCGCGCGCGCG 414
Qy  520 CAAAAAAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 577
Db  415 GSASAAAAAAGAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
Qy  578 GLGSGGTSGRGLGCGAGAAAAAAGAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCG 637
Db  469 P-AQCGPSGCG-----IAASASAGGCGGCG-----PAQCGPAGYCGCGSAV 507
Qy  638 AAAAAAGAGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
```

```

Db          508 AASAGA-----GSAGYGPQSQAS 525

RESULT 7
US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078888
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-05-28

```

;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 24.4%; Score 863.5; DB 12; Length 1894;
Best Local Similarity 30.9%; Pred. No. 1e-44;
Matches 325; Conservative 30; Mismatches 303; Indels 393; Gaps 35;

Qy 5 TGGQMGGRGSMASGRGGLGGGAGAAAAAAGGAGGGGG--LGSQGTSG----- 56
Db 580 TGGAAAG--GTTACAGGGGATACAGTATATACATGTTCCGACTGTCAGAACTAGACATTG 638
Qy 57 -RGGLGGGAG-----AAAAAAG--GAGGG----- 83
Db 639 ATGATGACGAAGACCACATATCCACACAAATATAAAGCTGTGAGGTTTGTATTATC 698

Qy 84 -----YGGLSQGTSGRGLGGGAGAAAAAAGGAGGGGG 124
Db 699 CTTTGTGATGACATCATTTCAAGGGAATTTAAAGGCTGAAAAAAGAGAAACAGAGGAGG 758
Qy 125 -----LGSQGTSGRGLGGGAGAAAAAAGGAGGGGG 159
Db 759 AAGTAAAGAAATTTAAACCCAAAGGCACAAAAAATTTTAGTTTACTTCTATTGTGAGAGG 818
Qy 160 YGGL-----GSGTSGPGG-----YGPQQTSGRGLGGGAG-----AAAAAA 199
Db 819 AAGCTGAGGAAGAGAGGAGGAGGAAATTAATCGAGTTAGTCAGAGCATGAAGGCCAAAGCA 878
Qy 200 AAGGAGGGGG-----GLSQGTSGRGLGGGAGAA 232
Db 879 AAGTAGTCATGACTTGTCTTAAGGATGATCCACATCTCAGTTCTGTCTCCAGTTGTAGAAA 938
Qy 233 AAAAAAAGGAG-----OGGYGGGLGSQGTSG--RGLGGGAGAAAA 272
Db 939 GTGAAAAAGGTGATGCACCATTTAGTTGATGATGAGGAGGATGAAAGTGCAGAGCATG 998
Qy 273 AAAAAAGGAGGGYGGGLGSQGTSGRG-----GLGGGAGAAAA--AAA 312
Db 999 ATGAATATATTTGATGCTGATGAAAAAGACCTGATGAGAGAAAGATTTCCCAAAAAATTA 1058
Qy 313 AAAAGGA-----GQGYGGGLGSQGTSGPGGYG----- 340
Db 1059 AAAGGACACAAGTCGGAATGTTAAATCAGCTGGGAGAGGAGAGCTGCAGAGAAATCAG 1118
Qy 341 -----GQGTSGRGLGGGAGAAAAAAGGAG-----OGGYGGGLGSQGTSGRGGG 389
Db 1119 TCAGCCGAGTGAAGAGCTCAGAAAGGAGCAAGCAATTAANAAGCG--GNACTCTTAGCA 1177
Qy 390 GOGAAAAAAGGAGGGYGGGLG-----SQGTSGRGLGGGAGAAAAAAGGAG 442
Db 1178 GCAAAACAAAAAAGTAGAAAAATCCACCAAAACAGCAGAAAAAGAAAGTGAAGAGGAA 1237
Qy 443 GGAG-----OGGYGGGLGSQGTSGRGG-----LGGGAGAAAAAAGGAG----- 483
Db 1238 GAAGCCCTCCAGATGTTGCTGTTCCGCAATACAGAGAGAAAGCAAAAGTATCAAGCT 1297
Qy 484 -GGGYGG-----LGSQGTSGPGGYGPGQGTSGRGLGGGA----- 519
Db 1298 TTGAGGAAGCAACAGTCAAGAGAGGGAACCTTCCCGG--GAAGATCAGACCCCTTGCACTGCT 1356
Qy 520 -----GAAAA----- 524
Db 1357 GAACCATTTAACTCACTCAAGCAATTTCTGTAACACCTGAAATGACATTC 1416
Qy 525 -AAAAAGGAG--OGGYGGGLGSQGTSGRGLGG-----OGAGAAAA 562
Db 1417 TCAACAGAACTAGAAAGATGATGAAG--GATGGATGTCACATGCTACTTCACTTTGAGGATA 1475
Qy 563 AAAAAAGGAGGGYGGGLGSQ-----TSRGG----- 589
Db 1476 AAGCAGAAAGTGAAGATGCAAGCTGCAAGCTCAGATACATTTGAAATCTATGATC 1535
Qy 590 --LGG-----OGAGAAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAAAAA 642
Db 1536 CTCGAAATCCAGTGAATAAAGAGAGGAGGAGAAAGCAAAAGCTGATGAGAGAGAA 1595
Qy 643 AAGGAGGGYGGGLGSQGTSGPGGYGPGQOTS 673
Db 1596 AAGAAA-----GAAGATAAAATGAGAATA 1619

RESULT 8
US-09-804-898-2
; Sequence 2, Application US/09804898
; Patent No. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: DURING, MATTHEW
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS

;; PRIOR FILING DATE: 1997-12-17
;; PRIOR APPLICATION NUMBER: 60/068017
;; PRIOR FILING DATE: 1997-12-18
;; PRIOR APPLICATION NUMBER: 60/077450
;; PRIOR FILING DATE: 1998-03-10
;; PRIOR APPLICATION NUMBER: 60/077632
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077649
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/078886
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078939
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079664
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079786
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/080107
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080194
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080327
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080333
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/081049
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081070
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081195
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081838
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082568
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082569
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082704
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082797
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086023
;; PRIOR FILING DATE: 1998-05-18

;; PRIOR APPLICATION NUMBER: 60/086392
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086486
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087098
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087208
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087609
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 19.48; Score 687.5; DB 12; Length 1300;
Best Local Similarity 28.3%; Pred. No. 2e-34;
Matches 296; Conservative 17; Mismatches 341; Indels 391; Gaps 45;

```
QY 2 ASMTGGQMGKSGWASGRGLGG--QGAGAAAAAAGAGAGQ---GGYGLGSQGT--- 54
Db 123 AGATGGCTCCCTGCAGCGAGGACTGGGCCCTCAACATGGAGATCGGCACATCATCA 182
QY 55 SGRGGLGGGAGAGAAAAAGAGGAG---GGYGLGSQGTSGRG---GLGGGAG 104
Db 183 CGAGACGGAGGAGGTCCCAAGATCCCTCCGAGCAGTAAGAAGAGAGATCGTGGGGA 242
QY 105 AAAAAAAGAGAGGGYGLGSQGT-----TSRGGG-LGGGAGAGAAAAA 150
Db 243 TAAGAACTTCCAGAGGTGATGTGCTCTCACAGTCTTAGAAACCTGTGTCAAGAACTG 302
QY 151 AAGGAG-----QGYGGL-----GSGGTSRGGYGPQGTSGRGLGG 189
Db 303 CGGGCACCCGTTCCACGTCTGTGTGCCAGCCAGGACTCGTGGAGAGTGTGCTGTGAG 362
QY 190 G-----AGAAAA-----AAAAA----- 203
Db 363 GACCATCTGCCCAAGAACACCCACCACCATCGTGCATGACAAAGTGTCAACCTCAT 422
QY 204 -----GAGGG-----YGLGSQGT----- 218
Db 423 CCAGTCTGGGTGACGCGTTCGCGACGTCGCCGATCTGACAGGTGTGCTACCATCTA 482
QY 219 SGRGGLGGGAGAA--AAAAAAGAG-----QGYGLGSQGT--- 256
Db 483 TGAGGACCTCGGAGAGAAAGGCTGGAGTTCCCATGACTGACCTGGACATGCTGTCAACC 542
QY 257 -----SRGG-LGGG-----AGA-AAAAAAGAGGGY----- 286
Db 543 CATCCACACCCAGAGGACCGTGTCAACTCAGACACATCAGACAGGATTCGTG 602
QY 287 GGLGSQGT-----TSRGG----- 299
Db 603 GGCACCTGACTCCAGCAGCAAGGAGACTCTGGCCAGCAGTCTGCCCTCTGCCGCCCG 662
QY 300 -----LGGG-----AGAAAAAAGAGAG--QGYGLGSQGTSGPG 337
Db 663 CCATACTCTCCGGTGACACGCCCATAGCACCACCCGGAACAGATGGGAGCTGCCG 722
QY 338 YPGQOQ--TSRGGGLGGAGAAAAAAGAGGGYGLGSQGTSG--RGLGGGQ--A 393
Db 723 AGTGACCTGGAGTGTGATGGGACGTGAGGTGATCTCGAGATGCTGACGGAGCTG 782
QY 394 GAAAAAAGAG-----AGGGYGLGSQGTSGRGLGGGAGAAAAAAG 443
Db 783 GTGCCACCCAGCCGAGCCCGCAGACCTGGAGCTGCTGAG-----GAGCTCAACGCCAC 838
QY 444 GAGGGYG--GLGSQTSRGG--LGGGAG-----AAAAAAG 481
Db 839 GTGCCAGCCATGCACGCGGGTCTGTAGTGATACCTGTCCCGGGCCCATGCCCAAG 898
QY 482 GAG--QGYGGLG-----SQTSRGGYGPQGTSGRGLGG-----QGAGAAA 524
Db 899 GAGCCCTTCAGAGCCACACATGCCATCGAGGCTTGCTGGAGGCTGCCACAGTGGNA 958
QY 525 AAAAAAGAG-----QGYG-----GLGSQTSRGL 552
Db 959 TTCTGCCAGCCTATGTCTCCCTACCTGTCTGTGATACCTGTCCCGGGCCCATGGCTTG 1017
QY 553 GG-----QGAGAAAAAAGAGAGGGYGLGS-----QTSRGLGGGAGAAAAA 603
Db 1018 GGCCTAGAGGTAGGTTGAGGTGTGGAGGCCCTTGAGAGGCTCGCGGGCCCGAGG 1077
QY 604 AAGAGAGGGY-----GLGSQTSRGLGGGAGAAAAA-----A 644
Db 1078 TAGGAGCTGCACTCTGCCGCACTG-----GGCAGATCTCATCAGCCCGGCTGCA 1132
QY 645 GAGGAG-----YGLGSQTSRGG 665
Db 1133 GGTGAGGCTTCAGGGGATGCTGGG 1157
```

```
RESULT 10
US-09-060-854B-2
; Sequence 2, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060,854B
; CURRENT FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: B. amyloliquefaciens
US-09-060-854B-2

Query Match 19.1%; Score 675.5; DB 10; Length 1497;
Best Local Similarity 30.7%; Pred. No. 1.1e-33;
Matches 265; Conservative 22; Mismatches 330; Indels 245; Gaps 28;

QY 27 AGAAAAAAGAGAGGGYGLGSQGTSGRGLGGGAGAAAAAAGGA----- 79
Db 70 AATGAAAAAAGAGAG-----GATAAAGAGTCAGAGGCCAAAAAGTATGGATCAGTTT 124
QY 80 -----GGGYGGLGSO--GTSG-----RGL 98
Db 125 GCTGTTTGTATTAGCGTTAATCTTTACGATGGCGTTCGGCAGCACATCTCTGCCCAGGC 184
QY 99 GGGAGAAAAAAGAGAGGGYGLGSQGTSGRGLGGGAGAAAAAAGAG-- 156
Db 185 GCGAGGAAATCAACCGGGGAAA-AGAAATATATTGTCGGGTTTAAACACACATGAGCA 243
QY 157 QGYG--GLGSQTSRGGYGPQGTSG-----RGLGGGAGAAAAAAGGA 205
Db 244 CGATGAGCCCGGTAAGAAGAGATGTCTTTGAAAAAGCGGGGAAAGTGCAAAAGC 303
QY 206 GGGYGLGSQTSRGGGLGG--QGAGAAAAAAGAGAGGGYGLGSQTSRGG 263
Db 304 AATTCAAATATGTAGACGAGCTTCACTCAATTAACGAA-----AAGCTGTA--AAG 357
QY 264 GGGAGAAAAAAGAGAGGGYGLGSQGTSGRGLGGGAGAAAAAAGAGGAG 323
Db 358 AATTGAAAAAGACCAGCGCTGCTTACGTTGAAGAG--ATCAGTAGCACATGCGT 414
QY 324 YGLGSQTSRGGYGPQGTSGRGLGGGAGAAAAA----- 364
Db 415 AGCCGAGTC-----CGTGCTTACGCGGTATCAAAATTAAGCCCTGCTCTGCACCTCT 470
QY 365 -AAGG-----AGGGYGLGSQTSRGGGAGAAAAAAGAGAG----- 408
Db 471 CRAAGGTACACTGGATCAATGTAAAGTAGCGGTTATCGACAGCGGTATGCTGATTCTCT 530
QY 409 -----QGYGGLGSQTSR--RGLGGG----- 430
Db 531 CATCTCTGATTAAAGGTAGCAGCGAGCGAGCATGTTCTTCTTGAACAAATCTCTTC 590
QY 431 --AGAAAAA-----AAGGAG-----QGYGGLGSQTSRGGGAGAAAAA 477
Db 591 CRAAGAACAACTCTCACGGAATCACTGTTGCCGACACAGTTTCGCGCTCTTAATAACTCA 650
QY 478 AAGAGAG-----QGYG-----GLGSQTSRGG-----YGPQQT 508
Db 651 ATCGGTGTATTAGCGCTTGGCCAGCGCATCACTTTACGCTGTAAAGTTCTCGGTGCT 710
QY 509 SGRGGLGGGAGAAAAAAGAGAGGGYGLGSQTSRGGGAGGAG--GAAAAA 566
Db 711 GACGGTTCCGGCAATACAGCTGGATCAATACGGAATCGAGTGGCGGATCGCAACAT 770
```

QY 567 AAGGA-----GQGGYG-----LGSQGTSGRGLGSGQ 594
Db 771 ATGCAGCTTATTAACATGAGCCTCGCGGACCTTCTGTCTCTGCTGCTTTAAAGCGCA 830
QY 595 AGAAAAAAGAGAG-----GGYGLGSQGTSGRGLGSGQAGAAAAAAGG-----646
Db 831 GTTGATAAGCGGTTGCATCGCGCTCGTAGTGTGTCGGCAGCGGTAACGAAGGCAC 890
QY 647 -----AGQGGYGGLGSQGTSG 662
Db 891 TCCGCGACCTCAAGCACAGTGS 912

RESULT 11
US-10-029-217A-24
; Sequence 24, Application US/10029217A
; Patent No. US20020164735A1
; GENERAL INFORMATION:
; APPLICANT: OLSON, ERIC N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO A CARDIAC-SPECIFIC
; FILE REFERENCE: US/SD:695US
; CURRENT APPLICATION NUMBER: US/10/029,217A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/257,761
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-029-217A-24

Query Match 19.08; Score 672; DB 9; Length 3907;
Best Local Similarity 30.94; Pred. No. 3.7e-33;
Matches 258; Conservative 15; Mismatches 342; Indels 220; Gaps 33;

QY 26 CAGAAAAAAGAGAGAGGQGG---YGLGSQGTSGRGG-----LGGGAG-----66
Db 886 GCCAAGTACAGCGGAGGCCCTGG-CAAGCAGCGGGAGCCCCCAGTAGCAGCCTCT 944
QY 67 -AAAAAAGAGAG-----QGYY-----GGLGSQGTSGRGLGSGQAGAAAA 108
Db 945 CCACTACCAATAGCAGCTCCAGCTCGGCGCCCTGGGCCCTGTGG-GCTGGCAGTCA 1003
QY 109 AA-----AAAGAGAGGYYGLGSQGTSGRGLGSGQAGAAAAAAGGAG 156
Db 1004 AACAGCACCTCACTGACTGGCAAGCGG--GGAGCCCTGCGCGCAACCTGGACGACATGA 1061
QY 157 QGGYGLGSQGTSGPGYGPQGTSGRGLGSGQAGAAAAAAGGAGGQGG-----CAAAAAA 201
Db 1062 AGTGGCAGACTGAGCAGAGAGCTGAAGTTGCGATCACTCCCTGTCTCGGGCACCANA 1121
QY 202 AGGAGQGGY-----GGLGSQGTSGRGLGSGQAGAAAAAAGGAGGQGG-----YGLGSQ 255
Db 1122 CTGAGCTGATTGAGCGCCTTCAGCCTATCAAGACCAATCAGCCCTGTGCGCAGGAGCC 1181
QY 256 TSGRGG-----LGGGAG-----AAAAAAGAGAGGQGGYGLGSQGTSG-----296
Db 1182 CCAAGGCCCTGCGGCCACCTCTATCTGCAACAAGGCTGGCGAGGTGGTG--GTAGCCTT 1239
QY 297 --RGLG--QGAAGAAAAAAGAGAGQGGYGLGS-----QGTSGP 335
Db 1240 CCAAGGCCCGGCTGAGCAGCGGCCAGCCCTGGTGGCAGCAGCCCTGGCTCCAGCTGA 1299
QY 336 GGYGPQGTSGRGLGSGQAGAAAAAAGAGAGQGGYGLGSQGTSGRGLGSGQAG-394
Db 1300 GG-----TGTTGGTGGCCACGGTGGCCAGCAGTGGGTGTGAAGTT-TGGCAGCAGG 1352
QY 395 -----AAAAAAGAGAGQGG---CYGLGSQGTSGRGLGSGQ 430

Db 1353 GCTCCACGCCCCCGCTGTCTCCACCCCTCGGAGCGCTCACTCTCAGCAGGGCGATG 1412
QY 431 AGAA-----AAAAAAGAGAGGQGGYGLGSQGTSG-----RGGLGGGAGAA 472
Db 1413 AAACCTCCACCCCGGACACCTTTGGTGAGTGTGACATCACTCTGACGACAGTGA 1472
QY 473 AAAAAAAGGAGQGG-----GYGG-----LGSQGTSGPGYGPQGTSGRGLGSGQ-----518
Db 1473 CCCTGCGCCTCGCAGCTGCAGATCCTCTGAAGGAGGAGGCCCCCGGCGGCTCT 1532
QY 519 --AGAAAAAAGAGAGGQGGYGLGSQGTSGRGLGSGQAGAAAAAAGAGAGGQGG 575
Db 1533 GTTGCTGAGCCCTGGGGGCGGCGAGCTAGAGG-GCGCGACAGGACCATGCTG 1591
QY 576 YGGLGSQGTSGRGLGSGQAGAAAAAAGGA--GQGYGGLG-----SQTSGRGG 627
Db 1592 CAG-----GAGAAAGACAAGCAGATCGAGCGGCTGACCGCATGCTCCGG 1636
QY 628 LGGAGAGAAAAAAGAGGQGG-----YGLGSQGTSGPGYGP 669
Db 1637 CAGAAGCAGCAGCTGCTGAGCGGCTCAAGCTGAGCTGGAGCAGGAGAACGGAG 1691

RESULT 12
US-10-096-961-1
; Sequence 1, Application US/10096961
; Patent No. US20020155572A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000849DIV
; CURRENT APPLICATION NUMBER: US/10/096,961
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/232,632
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/738,884
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-096-961-1

Query Match 18.74; Score 662; DB 9; Length 2211;
Best Local Similarity 27.34; Pred. No. 9.6e-33;
Matches 281; Conservative 22; Mismatches 355; Indels 372; Gaps 37;

QY 2 ASMTGGQOMGRGSMASGRG-----GLGGGAGAAAA-----AAAAAG-----77
Db 431 ACCAGGACAGCAAGTAGCTTCAAGGAGATCAAGAGCCTCTGAGATGCTCAACCTG 490
QY 46 YGGLGSQ-GTSGRG-----GLGGGAGAAAA-----AAAAAG-----77
Db 491 ACATGAACGACATGTACCCCTACCTCTTCAAGGAGTGTGACCCTCCCAACACGACC 550
QY 78 -----GAGQGG-----CYGLGSQGTSGRGLGSGQAGAAAAAAGGAG-OGG 121
Db 551 GTCTAGAGGGGCTGAGATCGAGGAGTTCTCTGCGCGGCTGCTGAAGGCGCGGAGCTG 610
QY 122 YGGLGS-----QGTSGRGLGSGQAGAAAAAAGAGAGQGG-----GYGLGSQ 167
Db 611 AGGAGATCTTCATCAGTACTCG--GGCAGGACCGCGTGTGAGTGGCCCTGAGCTGCTG 669
QY 168 TSG-----PGYGPQGTSGRGLGSGQAGAAAAAAG-----203
Db 670 GAGTTCTTGAGGAGCAGGCGCAGAGGCCACACTGGCCCGCCAGCAGCTCAT 729
QY 204 -----GAGQGGYGLGSQGTSG--RGGLGGGAGAAAAAAGGAGQ 246

Db 730 CAGACCTATGAGCTCAACGAGACAGCAACGACATGAGCTGATGACACTGGATGGCTTC 789
QY 247 GYGGLGS---QGTSGR---GGLGGOGA-----GAAAAAAGAGGAGGCGGGLGS 291
Db 790 ATGATGTAAGTGTGTTCGGGAGGGGAGCTGCTTGGACAACACCCACACGT-----GTGT 845
QY 292 QGTSRGLGGGAGAGAA-----AAAAA----- 315
Db 846 GTTCCAGGACATGAACACAGCCCTTCCCACTACTTCACTCTCTCTCTCCCAACACCTA 905
QY 316 -----AGAGGCGGGLGSQGTSPGGYGPQQTSRGLGGGAGGAGAAAA 361
Db 906 TCTGACTGACTCCAGATCGGGGGGCCAGCAGCAGCAGCGCTATGTAGGCGCTTTC 965
QY 362 AAAAAAGAGGCGGGLGSQGTSGRGLGGGAGAAAAAAGAGGAGGCGY----- 413
Db 966 CAGGATGCGCTGAGCTGAGCTGCTGGGAGGGCCAGAGGGGCGCCCTCAT 1025
QY 414 -----GLGSQGTSG---RGLGGGAG 432
Db 1026 CTATCATGCCATPACCTCACCTCCAAGATTCTTCCGGGAGCTGCCCAAGCCCTGCG 1085
QY 433 -----AAAAAAGAGGAG 447
Db 1086 CGACCATGCTTCACGCTGTCCCTTACCTGTCTATCTCTATCTCTGAGAACCACTGCGG 1145
QY 448 GYGGLGSQGTSGRGLGGOGA-----GAAAAAAGAGG---GGY 489
Db 1146 GCTGGAGCAGCA--GGCTGCCATGCGCCGCCACCTCTGACCATCTCTGGGGGACATGCTG 1203
QY 490 GLGSQGTSPGGYGPQQT-----SGRGLGGG-----AGAAAAAAGAG 531
Db 1204 GTGACACAGCGCTGACTCCCAATCCGAGGAGCTGCCATCCCAAGAGCTGAAG 1263
QY 532 GAGGCGGGLGSQGTSGRGLGGGAGAAAAA-----AAAAAGAGGCGGGLGS 581
Db 1264 GCGCGG-----GTCTGTGTAAGGGAAGAGCTGCCCGCTGCTCGGAGCGAGGATGG 1316
QY 582 -----QGTSG---RGLGGGAGAAAAAAGAGGCGYGG--LGSQGTSGRGL 628
Db 1317 CCGGGCTCTGTGGATCGGGAGGGGGAGGAGGATGACGAGGAGAGAGAGAGGT 1376
QY 629 GGAGAGAAAAAAGAGGAGG-----YGLGSQGTSG-----PG 664
Db 1377 GGAG--CTGAGCGCAGAGCGGCTGCCCAAGCAGATCTCCCGGAGCTGTCCGCGCTCG 1435
QY 665 GYGPQQTS 674
Db 1436 CTGTGTACTG 1445

RESULT 13

US-09-879-957-37

; Sequence 37, Application US/09879957

; Patent No. US20020034755A1

; GENERAL INFORMATION:

; APPLICANT: SPARKS, Andrew B.

; HOFFMAN, No. US20020034755A1h

; KAY, Brian K.

; FOWLES, Dana M.

; MCCONNELL, Stephen J.

; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

; USING SAME

; NUMBER OF SEQUENCES: 227

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

;

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-879-957-37

Query Match 18.4%; Score 652; DB 10; Length 1400;

Best Local Similarity 28.3%; Pred. No. 2.7e-32;

Matches 274; Conservative 24; Mismatches 325; Indels 344; Gaps 32;

QY 16 ASRGGGLGGGAGAAAAAAGAGGAGGCGGGLGSQGTSGRGLGGGAGAAAAAAGAG 75

Db 2 AAAGGAGGAGAGTGTCAAAA--GAAGATGCGA-----GGAAAAAGCAACA 50

QY 76 AGAGGGYGLGSQGTSGRGLGG-----QGAGAAAAAAGAGGAG 120

Db 51 GGAAGCACAAGACAGCTGGTGGCTTTTCATCAACACCAAGAACCACTAAGCA-- 108

QY 121 GYGGLGSQGTSGRGG-----LGGQ-----GAGAAAA----- 147

Db 109 -----GCTGTCAGGACCCCTGCTCCACTCCAGAAAAAGGTCCACTTACCTTTCTGCA 163

QY 148 -----AAAAAGAGGCGGGLGSQGTSGRGG-----YGPQQTSRGG---L 186

Db 164 CAGGAAAAATGTAAGTGTGTATTACCGGCGCACTGTACCCCTTTGTAATCCAGAGCCAT 223

QY 187 GGAGAAAAAAGAGGAGGCGGGLGSQGTSGRGLGGGAGAAAAAAGAGGAG 246

Db 224 GATGAATCACTATCAGCCAGGAGACATAGTCATGGTGGATGAAAGCAAACTGGAGAA 283

QY 247 GYGGLGSQGTSGRGLGGGAGAAAAAAGAGGAGGCGGGLGSQGT-----SGRG 298

Db 284 CCGG--CTGCTGGAGGAGAAATTAAGGAAGACAGG---GTGTTCCCTTGCNACT 338

QY 299 GLGGGAGAAAAAAGAGGCGGGLGSQGTSGRGGYGPQQTSRGLGGGAGGAGAA 358

Db 339 ATCAGAGAAAAATCCCAAAAAATGAGTTCCTCCGCTCCAGTGAACCACTGACTGATTCAA 398

QY 359 AA-----AAAAAGAGGCGG--GLGS-----OGT----- 382

Db 399 CATCTCCCTTCCCAAACTGGCTTCGTCAGACCCCGCCCTTTTGGCAGTAACCT 458

QY 383 -----SGRG-----GLGG----- 390

Db 459 CTCAGAGCCCTCCAGACCCCTTAATACTGGCCGCACTCAGCTCCAGCTGCCCAACA 518

QY 391 -----QGAGAAAAAAGAGGCGGGLGSQGTSGRGLGG---QGAGAAAAAAG 440

Db 519 GCAGGAATGAGAAACCAAGAACGATAACTGGGATGATGGGCGGCGCCCTCTCTCA 578

Qy 441 AAG-----GAGGGYGGGSGQTSRGGGCGG-----AGAAAAA--- 479
Db 579 CCGTTCCAAGTGC CGGC-CAGTTAAGCAGAGGTC CGCCTTTACTCCAGCCACGGCCACT 637
Qy 480 -----AGGAGGGYGGGSGQTSRGGGCGG-----AGAAAAA--- 513
Db 638 GGCTCTCCCGTCTCTGCTAGGCGAG--GGTGAAGAGGTGAGGGGCTACAAGCTC 695
Qy 514 -----LGCGAGAAAAA----- 530
Db 696 AAGCCCTATATCTTGGAGAGCCAAAAAGACACCACTTAATTTTAAACAAATGATG 755
Qy 531 -----GGAGGGYGGGSGQTSRGGGCGGAGAA---AAAAAAGAGAG 573
Db 756 TCATCACCGTCTCTGGAACAGACATGTGTGTTGGAGAGTTCAAGTTCAGAGG 815
Qy 574 GGYGGLG---SOGT-----SGRGLGGGAGAAAAAAGGAG 610
Db 816 GTTGGTCCCAAGTCTTACGTGAACATCTTTCAGGCCCATTAAGGAAGTCTACAAGCA 875
Qy 611 QGG----YGGGSGQTSRGGG---LGCGAGAAAAAAGGAG-----QGGYGGG 657
Db 876 TGGATTCGTGTTCTCAGAGAGTCTGCTAGTCTAAGCGAGTAGCCTCTCCAGCAGCCA 935
Qy 658 QGTSGP 664
Db 936 AGCGGAG 942

RESULT 14

US-09-756-071B-20
; Sequence 20, Application US/09756071B
; Patent No. US20020052307A1
; GENERAL INFORMATION:
; APPLICANT: Tryggevason, Karl
; Kallunki, Pekka
; Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
; STREET: 1100 Superior Ave, Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,071B
; FILING DATE: 08-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/663,147
; FILING DATE: 150-September 2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard, J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 20014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-861-5582
; TELEFAX: 216-241-1666
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-756-071B-20

Query Match 18.4%; Score 651.5; DB 10; Length 720;
Best Local Similarity 32.5%; Pred. No. 1.7e-32;
Matches 249; Conservative 17; Mismatches 314; Indels 187; Gaps 35;
Qy 16 ASRGRGLGGGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAA--- 75
Db 8 ATAGGAGGTTGGCCAGTCAATAGT-----TACTTTATGATTG-----CTAACCC 54
Qy 76 AGGAGGGYGGGSGQTSRGG---LGCGAGAAA-----AAAAAAGGAGGGYGG 124
Db 55 TGGTGAGCAGGAAGTATGTGGACAGAGAGAAACCCCTTGGTTCAGCCTGGAGAAA-GG 113
Qy 125 LGSQTSRGGGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGCCCTGTGTGACTCTC 172
Db 114 AGAGGTT-----GACCCCTAACTGGAG-GGTGGAGAGGAGCCCTGTGTGACTCTC 162
Qy 173 -GYGPGQTSG-----RGLGGGAGAAAAAAGAGGGYGGGSGQTSRGGG-L 224
Db 163 CGACTGACTTCTTCTTCTGATG---TCCTTTAAGCCGGA-----GCTGATTC 207
Qy 225 GGAGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAA-----AAAAAG 279
Db 208 GGGCTGCTGCTTATTTCTGAGTTAGCCCTCTTAAGATTGGCCCTCCAGTTTGAAGAG 267
Qy 280 CAGGGYGGGSGQTS-----SGRGLGGGQ---AGAAAAAAGAGGGYGGG 329
Db 268 GGGCGG-GCTGCTGTCTACCTCTGTGATCTGCCCTGGACCCCGGAGAGAGGAGG 326
Qy 330 QTSRGGYGGGSGQTSRGGGCGGAGAAAAAAGAGGGYGGGSGQTSRGGG-L 388
Db 327 -----GCTCCGGGGAATCTCGCACATTCAGGCAAGGCTCCCGGCGCAGCC 375
Qy 389 GGQAGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAAAGAGGGAGG 448
Db 376 TCTGTGCCACACCTTTGGCCCG---GGCCAGGT-----GTGCGCCCTCTCGCTGCCAG 426
Qy 449 YGGLG-SQTSRGGGCGGAGAAAAAAG---GAGGGYGG---GLGSGQTSG-- 498
Db 427 GGGGAGCGGGCGGTGCGGGGAGCGATTTCCAGCCCGGTTGTGCTGTGTGTTGTC 486
Qy 499 -----PGYGPQTSR-----GGLGGGAGAAA-----AAAAAAG-- 532
Db 487 TGCCTCTGGAGGGCTGGGCTCTCTCTTATTCACAGGTAGTCACACCTGAAACACAGGCT 546
Qy 533 -----AGGGYGGG--SQTSRGGGCGGAGAAAA-----AAAAAGGAGQ 573
Db 547 CTCCTCTGTCTAG-GACTGAGTCAGGTAGAGAGTCGATAAAACACCTGATCAAGGAAA 605
Qy 574 -GGYGGGSGQTSRGG-GLGGGCGAGAAAAAAGAGGGYGGGSGQTSRGGG 630
Db 606 AGGAAGCACAGCGGAGCGAGGTGAGAACCAACCAACCGAGGGCGCG-----GGCAG 658
Qy 631 QCA-----GAAAAAAGAGGGAGGG---YGGGSGQTSGP 664
Db 659 CGACCCCTGACGCGGAGACAGAGACTGAGCGCGCGCGGACCGCCCATG 705

RESULT 15

US-10-052-586-525
; Sequence 525, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 : Search time 20.4171 Seconds
(without alignments)
3253.588 Million cell updates/sec

Title: US-09-490-291-6
Perfect score: 3543
Sequence: 1 MASMTGGQQMGSGMASGRG.....TSGIRAPSTSFHHHHH 691

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2138	60.3	718	2 A36068	major ampullate fi
2	1386	39.1	2639	2 T31328	fibroin - Chinese
3	1281.5	36.2	1901	2 F70806	hypothetical glyci
4	1260.5	35.6	1489	2 D70807	hypothetical glyci
5	1242	35.1	1079	2 B70807	hypothetical glyci
6	1198.5	33.8	1306	2 A70934	hypothetical glyci
7	1196.5	33.8	1381	2 E70806	hypothetical glyci
8	1180	33.3	1660	2 A70869	hypothetical glyci
9	1158.5	32.7	1538	2 H70846	hypothetical glyci
10	1152	32.5	1329	2 E70917	hypothetical glyci
11	1151	32.5	853	2 A70896	hypothetical glyci
12	1138.5	32.1	627	2 A44112	spidroin 2, dragli
13	1114.5	31.5	749	2 A70812	hypothetical glyci
14	1110.5	31.3	914	2 H70987	hypothetical glyci
15	1083	30.6	882	2 B70812	hypothetical glyci
16	1049.5	29.6	778	2 F70963	hypothetical glyci
17	1038.5	29.3	767	2 E70895	hypothetical glyci
18	1032.5	29.1	837	2 E70835	hypothetical glyci
19	1032	29.1	937	2 D70835	hypothetical glyci
20	1014	28.6	783	2 E70824	hypothetical glyci
21	1007	28.4	731	2 C70974	hypothetical glyci
22	992	28.0	714	2 A70807	hypothetical glyci
23	989	27.9	741	2 G70917	hypothetical glyci
24	988.5	27.9	801	2 F70824	hypothetical glyci
25	977.5	27.6	1011	2 F70620	hypothetical glyci
26	976	27.5	860	1 EAMS	elastin precursor
27	939	26.5	864	1 EART	elastin precursor
28	937	26.4	923	1 E70820	hypothetical glyci
29	923.5	26.1	667	2 A70893	hypothetical glyci

30 903.5 25.5 694 2 F70868 hypothetical glyci
31 893 25.2 484 2 G70846 hypothetical glyci
32 869 24.5 603 2 A70770 hypothetical glyci
33 853.5 24.1 465 1 S01820 glycine-rich cell
34 848.5 23.9 1585 2 T31611 hypothetical prote
35 833.5 23.5 576 2 A70900 hypothetical glyci
36 817.5 23.1 584 2 G70804 hypothetical glyci
37 809 22.8 770 2 S59623 tropoelastin - she
38 808.5 22.8 618 2 A70989 hypothetical glyci
39 804 22.7 639 2 D70931 hypothetical glyci
40 801 22.6 606 2 H70816 hypothetical glyci
41 800 22.6 591 2 B70523 hypothetical glyci
42 786.5 22.2 747 1 EABO elastin precursor
43 777.5 21.9 588 2 F70971 hypothetical glyci
44 776.5 21.9 615 2 H70589 hypothetical glyci
45 763 21.5 641 1 Q0BE31 nuclear antigen EB

ALIGNMENTS

RESULT 1

A36068
major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)
C:Species: Nephila clavipes
C>Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997
C:Accession: A36068
R;Xu, M.; Lewis, R.V.
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990
A:Title: Structure of a protein superfiber: spider dragline silk.
A:Reference number: A36068; MUID:50384959; PMID:2402494
A:Accession: A36068
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-718 <XUA>
A:Cross-references: GB:M37137; NID:g159711; PID:g159712
A>Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue

Query Match 60.3% Score 2138; DB 2; Length 718;
Best Local Similarity 66.3%; Pred. No. 3.6e-113;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

Qy 25 QGAGAAAAAAGAGAGGCGGGLGSGQT--SCRGGLGCGAG-AAAAAAGAGGAG 81
Db 1 QGAG-----AAAAAGAGAGGCGGGLGCGAGGCGGAGGAGAGAGAGAGAGGAG 56
Qy 82 GGYGGLGSGQTSRGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
Db 57 GGYGGLGSGQ-----AGRGGCGAG---AAAAAGAGAGGCGGCGGGLGSGQ-AGRGGLGCGGA 107
Qy 142 GAAAAAAGAGAGGCGGGLGSGQTSRGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAG 201
Db 108 G-----AAAAAGAGAGGCGGGLGSGQ-----AGRGGCGG-----AAAA 142
Qy 202 AGAGAGGCGGGLGSGQTSRGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252
Db 143 AGAGAGGCGGGLGSGQ-AGRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
Qy 253 SQGTSRGRGLGCGAG 291
Db 202 SQG-AGRGGLGCGAG 260
Qy 292 QGTSRGRGLGCGAG 350
Db 261 QG-----AGRGGEGAG-----AAAAAGAGAGGCGGGLGSGQ-AGGCGGGLGSGQAGRGL 311
Qy 351 GCGAG 392
Db 312 GCGAG 370
Qy 393 AGAAAAAAGAGAGAGGCGGGLGSGQTSRGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAG 452
Db 371 AG---AVAAAAGAGAGGCGGGLGSGQ-----AGRGGCGAG-----AAAAAGAGAGAGAG 419

```
QY 453 LGSQGTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQT-----SGP 499
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 LGNQG-AGRGLGGOGAG---AAAAAGAGAGGGYGGGLGNQAGRGOGAAAAAGGAGQ 475
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 500 GGYGP-GQOTSGRGLGGOGAGAAAAA-----ACGAGGGYGGGLGSGQTSRGL 552
||||| ||: || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 GYGGGLGSGQAGR---GGCAGAAAAAVCAGQEGIRGQAGGGYGGGLGSGQ-SGRGL 531
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 GQGGAGAAAA-----AAAAAGGAGGGYGGGLGSGQTSRGLGAG 592
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 GQGGAGAAAAAGGAGGAGGGYGGGLGSGQAGAAAAAGVQGGYGGGLGSGQ-----AGRG 587
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 593 QGAGAAAAAAGAGAGGGYGGGLGSGQTSRGLGGOGAGAAAAAAGAGAGGGY 652
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 QGAG-----AAAAAGAGGGYGGGLGSGQY-GRGLGGQAG-----AAAAGGAGGGY 636
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 653 GGLGSGQTSRGGYGPQQTSGIRIRAPSTS 683
||:||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 637 GGVGSGASAA-----SAAASRLSSPOAS 659
||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
T31328
fibroin - Chinese oak silkworm
C:Species: Antherea pernyi (Chinese oak silkworm)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A:Reference number: 220995
A:Accession: T31328
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
A:Genetics:
C:Introns: 14/3

Query Match 39.1%; Score 1386; DB 2; Length 2639;
Best Local Similarity 40.6%, Pred. No. 1e-70;
Matches 400; Conservative 23; Mismatches 248; Indels 315; Gaps 33;

QY 5 TGGQMGGRSMASG-RGGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQ-----52
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 SGSSNNAAGSAGAGGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAA 210
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 53 -----GTSRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSRGLGGQAG 104
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 AAAGSGAGSGGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAA 270
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 105 AAAAAAAGAGAGGGYGGGLGSGQTS-----GRGGLG 137
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 AAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAGGGYGGYGGY 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 138 GCGAGAAAAA-----AGGAGGGYGGGLGSGQ-----166
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 SDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAAAA 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 167 -----GTSRGGYGPQQTSGRGLGGOGAGAAAAAAGAGAGGGYGG-----212
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 391 AAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAGGGYGGYSDS 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 -----LGSQTSRGLGGOGAGAAAAAAGAGAGGGYGGGLG 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 GYVSDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAGAGAGAG--GS 507
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 254 QGTSRGLGGOGAGAAAAA-----AGGAGGGYGGGLGSGQT-----SGR 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 YGW-GDGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAGAGAG 566
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 298 GGLGGOGAGAAAAAAGAGAG--GGY-----GGLGSGQT-----SGPGY 338
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 567 DSAAGSAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAGAGAG 626
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 339 GPCQQTSGRGLGGOGAGAAAAA-----GGAGGGYGGGLGSGQT-----S 383
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 627 GGGYGW-GDGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAG 685
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 384 GRGGLG-----QGAGAAAAA-----ACGAGGGYGGGLGSGQT-----420
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 686 GAGYGGYGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAG 745
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 -----SGRGLGGOGAGAAAAAAGAGAG--GGY-----GGLGSGQT-----458
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 746 AARRAGHDSAAAGSAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAG 805
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 459 -SGRGLG-----QGAGAAAAAAGAGAGGGYGGGLGSGQT-----496
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 806 GSGAGAGGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAG 865
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 -----SGPGYGPQQTSGRGLGGOGAGAAAAA-----AGGAGGGYGGGLGSGQT- 546
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 866 AAAAGSAGYGGYGGYGGYSDSAAAAAAGAGAGGGYGGYSDSAAAAA 924
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 -----SGRGLGGOGAGAAAAAAGAGAG--GGYGGYGGGLGSGQTSRGLG 593
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 925 AAAAAAARRAGHDSAAAGSAAAAAAGAGAGGGYGGYSDSAAAAAAGAGAG 977
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 594 GAGAAAAAAGAGAGGGYGG-----LGSQGTSGRGG 627
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 978 SAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAGGGYGGYSDS 1037
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 628 LGGCGAGAAAAAAGAGAGGGY-----GGLGSGQ-----GTSG 662
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1038 YGSDSAAAAAAGAGAGGGYGGYGGYSDSAAAAAAGAGAGGGYGG 1097
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 663 PGYGGPQQTSGIRIRAPSTSFEHHH 688
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1098 DGGYGGSSAAAAAARRAGH 1123
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
F70806
hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70806
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1901 <COL>
A:Cross-references: GB:AL020222; GB:AL123456; NID:g3261554; PIDN:CAAL17745.1; PID:g292
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3508
C:Superfamily: collagen alpha 1(IV) chain

Query Match 36.2%; Score 1281.5; DB 2; Length 1901;
Best Local Similarity 39.5%; Pred. No. 5.6e-65;
Matches 354; Conservative 28; Mismatches 273; Indels 241; Gaps 39;

QY 7 GQMGGRSMASGRGGLGGO-----GAGAAAAAAGAGAGGGYGGGLGSGQ-----52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 GQGGAGGAGSDGALGGTGGTGGGAGGAGRALLLGAGCGGGLGAGGGGGTGGAG 339
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 53 -----GTSRGLGG-----QGAGAAA-----AAAAAAGG-AGCGYGGGLGSGQ 90
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17749.1; PID:g292444
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv3512
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
Query Match 35.1%; Score 1242; DB 2; Length 1079;
Best Local Similarity 40.7%; Pred. No. 6.1e-63;
Matches 316; Conservative 34; Mismatches 306; Indels 120; Gaps 29;
QY 5 TGGQMGGRGSMASGRGLGCGGAGAAAAAAGCA-GGGYGG-----LGSGQ 53
DB 58 TGGQGGNGN--GGNGTGGKGTGGDGGALAGSSGAGGKGGGADGACAGTGSAPGTAG 115
QY 54 TSGRGLGGQ-----AGAAAAAAGCA-GGGYGG-----LGSGQ 101
DB 116 TGGDGGKGGGIGAGCTTGPVGTGASGCTGSGGAGGTGGDGAAGGTAGAGGAGNG 175
QY 102 -----GAGAAAAAAGGAGGAGG-----LGSGQTSRGLGCGGAGAAA 145
DB 176 KGGDGGAGVTSSTAGNSGAGGSGKGGACAGGAGATPGANGIAGNGDGD--GAAG 233
QY 146 AAAAAAGGAGCGGGLGSG-----GTSGPGYGPQQTSRGGGLGGGAGAAAAA 201
DB 234 AVGISGATGAGDGHGHTGNAAGNGTGGAGGSGTGVGGGTGGTGGNGGAIGCAGD 293
QY 202 AGGAG-CCGYGGLGSG-----GTSGRGLGGGQ--AGAAAAAAGGAG 244
DB 294 AGGSNSGNGGIGKGNAGAGGAGNGTGVANGTGGDGGNGAAGATAGSNGGAG 353
QY 245 QGYGGLGSGQTSRGLGSGQ-----AGAAAAAAGGAG-----QGGY 286
DB 354 TGSAGNG--GTGGRGSGGAGDGGGKGGAGDGEVGGAGGAGGSPNTPSGN 411
QY 287 GGLGSGQTSRGLGCGGAGAAAAAAGGAGCGGGLGSGQTSGP---GGYGPQQ 343
DB 412 GGQGGGSGGAG--GAAGAGGAGGANGTAGNGGQGGAGTGGAGAASSATNGSGGAGG 470
QY 344 TSGRGLGGGAGAAAAAAGGAGGCGGGLGSGQTSRGLGSGQ-----A 393
DB 471 TGGDGGCGAGTGGAGTGGAGDGGCGGQGGAGG--GAGGCGGAGGAGTGGCGNGTITG 529
QY 394 GAAAAAAGG--AGGCGGLGSGQ--GTSGRGLGCGGAGAAAAAAGGAGCGGY 450
DB 530 GTAGTAGAAGNGAAGKGGAGGCGGTGGTGGCGAGGCGGAGGTGGDRTVGGGTVPAGS 589
QY 451 GGLGSGQTSRGLGCGGAGAAAAAAGGAGGCGGGLG-----SQGTSPPGCGP 505
DB 590 GGQ--GNAGGGAGGQ--GGADGSGDGDGAGTGGNGNGNRNSNGTGGAGGNGG 645
QY 506 QOTSGRGLGGGAGAAA-AAAAAAGGAGGCGGY-GLGSGQTSRGLGCGGAGAAAA 563
DB 646 GAGGAGGAGSGGTGGNGAGGAGDAGAGNGNGNGTGGNGNGNGIAGMNGGAGT 705
QY 564 AAAAAAGGAGCGGGL-----GSQGT--SGRGLGCGGAGAAAAAAGGAGCGGGLG 618
DB 706 GSGNGGNGSGGNGAGMGGNSGTGSDGGAGGNGAAGTGGTGGDGLTGTGTGSGG 765
QY 619 SQGTSRGLGCGGAGAAAAAAGGAGGCGGYGGLGSGQTSRGGYGPQQTSG 674

DB 766 --GTGDCGNGNGANDNTANMTAAGDGGNGDGGFGGAGAGGGGLTAGANGTG 819
RESULT 6
A70934
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70934
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70934
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1306 <COL>
A:Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAAL17449.1; PID:g290
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0578c
C:Superfamily: collagen alpha 1(IV) chain
Query Match 33.8%; Score 1198.5; DB 2; Length 1306;
Best Local Similarity 38.4%; Pred. No. 1.9e-60;
Matches 323; Conservative 37; Mismatches 293; Indels 189; Gaps 34;
QY 11 GRGSMASGRGGLGCGGAGAAAAAAGG-----AGQGYGGLGSGQTSRGLGCGGAG 66
DB 399 GTTSTSGNGDGGKADAISSGQTGANGRGDGGQVNGGAG--GAGRGAGGLGFG 456
QY 67 AAAAAAAGG-AGQGYGG--LGSQTSRGLGCGG-----GAAAAA 114
DB 457 SEAPRPGAGGTGAGGNGGTQAGDGTGGAGGAGDGGSGGAGSIFGNASAPGAAGSP 516
QY 115 GG-----AGQGYGGL-----GSQTSRGLGCGG-----GAAA 146
DB 517 GNGNGGPGGAGGAGGAGGLAALASGONGSQGAGDGGGNGGTPNGHGCAAGALGV 576
QY 147 -----AAAAAGGAGCGGGLGSGQTSG--PGYG-PQ--QTSRGLGCGGAGAAAA 199
DB 577 NGVGAGGHHGDDPGVGGAGGSGSSTPGANGAPGNTPTSGNGNGRGRGADATGFGQT 636
QY 200 AAAG-----AGQGYGGLGSGQ-----GTSGRGLGCGGAGAAA 234
DB 637 GASGGRGDDGLVNGGAGGAGGNGSKGLPLGLRLGNPLDGGTGGNGAGGSGGAWAGN 696
QY 235 AAAAAAGG-AGQGYGGLGSGQ-----GTSGRGLGCGG--AGAAAAA 277
DB 697 GTTGAGGTGGVGTGGSGDGVNCSACADGHPGCTGGVGTGGKGGDGGDGAAPNGV 756
QY 278 AGGAGCGGGLGSGQTSRGLGCGGAGAAAAAAG--CAGCGCYGGLGSG--GTSQP 335
DB 757 AGSQPGGAGGDG--GTGVGNGGGRGIDGADGATAGARGDGGAGGAGGGRGTGTP 814
QY 336 GYGP-----GQTSRGLGCGG-----AGAAAAA 364
DB 815 GCAGPAGTGTGQAGGNGSGTGDPDGGNGANGSVFTNNGIAGNGNGNGSLRSGAG 874
QY 365 AAGGA-----GQGYGGLGSGQTSRGLGCGGAGAAAAAAGGA 407
DB 875 GSGGAGTGTGATGSSSTHVNCGNGNGNGNDHALUSGNAAGNGNGNGNSLRSGAG 934
QY 408 GQGYGGLGSGQTSRGLGCGGAGAAAAAAGAGGAGCGGYG-----LGSGQTS 460
DB 935 GHGNGGNASRGMGDGGTGGAGGNAGQINGGAGNGDGGTGGSDGNPALTGSGRGG 994
QY 461 RGLGCGGAGAAAAAAGGAGGCGGYGGLGSGQTS-----SGPGYGPQQTSG 509

A:Accession: E70917
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1329 <COL>
A:Cross-references: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g2131046
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1450c
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 32.5%; Score 1152; DB 2; Length 1329;
Best Local Similarity 35.8%; Pred. No. 7.6e-58;
Matches 326; Conservative 35; Mismatches 305; Indels 244; Gaps 35;

QY 7 GQMGRCMSASGRGLGGCGA-----GAAAAAAAAGGA----- 41
DB 133 CGAGGAGGIIWNGGAGGSGAPGVGGAGGAAGLFGTGGAGGAGGAGAGGAGCGSGWLL 192
QY 42 GGGYGGGLGSGQTSRGLGGGAGAAAAAAGGAGGGYGGGLGSGQTSRGLGGG 101
DB 193 GNGVGGAGGOSLLG-CATGGAGNAGLFGVGTGGPGPGPGVG--GTGGAGGLGTT 249
QY 102 GAGAAAAAAGG-AGGGYGGGLGSGQ-----GTSRGLGGGCGAGAAAAA 148
DB 250 LYAGGHHGAGGPGPIGGVGGHGGVGAAGLLGVGGHGGAGGAGCAAGAGGDLSPHG 309
QY 149 --AAAGCAGCGYGGGLG--QGTSGPGYGPQQTSGRGLG-----GQAGAAAA 196
DB 310 TSGVGGDAGDGTGGRGGLAGAGGAGGAGGVTGGTGGAGGAFSRALIVAGDNGDPGA 369
QY 197 AAAAAAGGAG--GGYGGGLGSGQTS-CRGLGGGAGAAAAAAGGA-----GQG 246
DB 370 GGAGTGGAGSTGAHGAAGASPTSGNGGAGGAGGNAHFSSGKAGGNGGAGGGLVNG 429
QY 247 GYGLGSGQT-----SRGLGGGCGAGAAAAAAGGA-CGGYGGGLG 291
DB 430 GAGGAGGAGPAPPSSGDPNGGGAGGAGGKGGDGAQAGDGGAGGKGGNGGA 489
QY 292 QGTSGRGL-----GGGAGAAAAAAGGA-----AGGA----- 319
DB 490 TGATGLNLGAGAGDGTGKGGNGAGGGGGGAGGGGKALAAATHQDSMGAGGAGNGGA 549
QY 320 -GGYGGGLGSGQT-----SGPGYGPQQTSGRGLGGGCA----- 355
DB 550 GGGGCGGCAKGTFDNGGDCVGGNGNGSGRTGGAGGIGGAGTAGACARGATPTSG 609
QY 356 -----GAAAAAAGGAGCGGYGGL-GSQGTSRGLGGGCGAGAAAAAAGG 406
DB 610 GNGGTGGGANATVAGGAGGAGGKGGNGLVNGGAGGKGGDMAGVAGSSPTTAGESGT 669
QY 407 AGQ-----GGYGGGLGQ-----GTSRGLGGGCGAGAAAAAAGG----- 444
DB 670 SQNGGAGGAGGAGGGRGDFGGDGTGGAGNGANGANATTPCAKGDGHHGPGAGGNN 729
QY 445 AGGGYGGGL-----GSGQTSRGLGGGCGAGAAAAAAGGAGCGGYG- 489
DB 730 GGGGPGGLAGNLFPQNGIQGVGSGKGGAGGLAGDGGNANGNPFAGDNGHGGNGG 789
QY 490 --GLGSGQTSRGGYGPQO-----TSGRGLGGGCGAGAAAAAAGGAGCGGYG 539
DB 790 NPGAGGGGSGGAGSTPCKAGHGTPTSGDGGDGGNGGN-----SQVYVNGGDDGNG 844
QY 540 GL-GSQGTSRGLGGGCGAGAAAAAAGG-----AAAGCAGCGYGGGLGSGQ 583
DB 845 GNGSAGTGGNGRGDGFAGGMSANATNPENGPNGNPGNGGAGGAGGAGLNG-GNGG 903
QY 584 TSGRGLGGGCGAGAAAAAAGGAGQ-----GGYGGGLG-----SQGTSGR 625
DB 904 AGNCGGLGFGGCAAGANGVAVAPQPGAGGHHGAGGNGGAGGNGGCVVSDGAGGA 963
QY 626 GGLGGG-----AGAAAAAAGGAGCGGYG-LGSQGTSGPGYGPQQTSGIR 676
DB 964 GGAGDGAFFAGDNGNGGCGAGAFAGGGGGRGDDGNGAGNAGGPGGTGTAGKAG-- 1021

QY 677 IRAPSTFEH 686
DB 1022 ---PAGSILH 1028

RESULT 11
A70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17207.1; PID:e125
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match 32.5%; Score 1151; DB 2; Length 853;
Best Local Similarity 40.4%; Pred. No. 6.3e-58;
Matches 301; Conservative 27; Mismatches 307; Indels 110; Gaps 28;

QY 7 GQMGRCMSASGRGLGGCGAGAAAA-----AAAAAGCAGCGGYGGGLGSGQTS----- 55
DB 131 GQDGGPGGLLVNGNGGTSTTAGVAGNGGAAGLNGGAGGAGGAGGAGGAGGAGGWL 190
QY 56 -GRGLGGGCGAGAAAAAAGGA-----GOGYGGGLGSGQTS-----RGLGGGGA 103
DB 191 YNGGAGGAGGTSVIPVAGGNGGAGGAGLWGTGCGAGDGGNGRSGPVNAGSAGNGG 250
QY 104 GAAAAAAGGAGGCGGYGGL-----SGQTSRGLGGGCGAGAAAAAAGGA 155
DB 251 AGGAAGLFGDAGAGNGGKGGAGGAATSFNTAGDGGAGGAGGAGGAGGAGGAGNG 310
QY 156 GGGYGGGL--SQGTSRGGYGPQQT-----GRGLGGGCA-----GAAAAAAGGA 205
DB 311 GSGGTGAGGAGTACAGGNGGAGGGGTGGLLFGNGGAGGHCAGAAAGNGLAAGNGVSSGGG 370
QY 206 GGGYGGGLGSGQTSRGLGGGQ-----AGAAAAAAGGAGCGGYGGL-----GSQ 254
DB 371 GAGTGGAGDG--GAGGAGNARLWVGAGGAGGAGDGGAGGKCGSLSGNANGAG 428
QY 255 GTSRGLGGGCGAGAAAAAAGGAGGCGYGGGLGSGQTSRGLGGGCGAGAAAAAAGGA 314
DB 429 GDSRGGTG--GAGGEGAGLLVGTGHHGDDGAG--GAATGDDGGAAGATGTAGAGG 484
QY 315 AAGCAGCGYGGGLGSGQTSRGGYGPQQTSGRGLGGGCGAGAAAAAAGG----- 368
DB 485 RGGAGGSGGSGDGGGAAGPAGW-----LFGDGGAGNGGCAAGGAGGAGGAGGNGG 539
QY 369 -AGGGYGGGLGSGQTS-----GRGLGGGCGAGAAAAAAGGA-----AAGCAGCGYGGGLGSGQ 418
DB 540 NNGNGGNGGNGGATGGWLYGNGGAGGCGATAGAGGAGANGVSTNGGTGGNGGIG-- 597
QY 419 GTSRGLGGGGA--GAAAAAAGGAG--OGYGGGL-----SQGTSRGLGGG--GAG 470
DB 598 GTGSGGAGGAGNAGLLVGGAGGAGGAGGAGGAGGTGFTSSDGGAGGDDGGGNGAG 657
QY 471 AAAAAAAGGAGCGGYGGGLGSGQTSRGGYGPQQTSGRGLGGGCGAGAAAAAAGGA 530
DB 658 GTGGLLFGAGGNGGPGSGGAGADIG--GNGAGNGGTGDNNGGNGGSGG--AGSGDG 712

QY 119 OGGYGLGSQTSRGLGCGAGAGAAAAAAGAGQGGYGLGSQTSRGGYG--- 175
Db 250 IGGAGAGGAATGTGGAGGAGSN-ALGLFLGLGSGGGSDSAMSGGAGGAGSGGAA 308
QY 176 -PQQTSGRGLGCGAGAGAAAAAAGGAG-----OGGYGGLGSQTSRGG 222
Db 309 SPFGIDIGIGGAGGHC-GAGTNGGAGGAGGAGSSSTVFALDLSWGGAGCGNGAATTGTG 367
QY 223 GLGGGAGAGAAAAAAGAGGAGGGYGLG--SQTSRGLGCGGAGAGAAA-----AAA 275
Db 368 GAG--GTGGFAVAPDFIGFAAYGAGGLGGAATGAGTGGTGGVAGGFAALGVGVGGA 425
QY 276 AAAGGA-----CGCGYGLGSQTSRGLGCGGAGAGAAAAAAG-----GAGQ 321
Db 426 GGAGGAATETGGTGGAGGLGVLLGGAGGAGGPG-GAASAGSGHGGTGGDALGLIGAGI 484
QY 322 GGYGGLG-----SQTSRGGYGPQQTSGRGLGCGGAGAGAAAAAAGGAGGGYGG 376
Db 485 GGVGVGGGAATDTGGNGGAGGSGTG-ILGVGAGAGHGGG--ASVGTGSGGAGGDPGF 541
QY 377 LGSQTSRGLG-----GQAGAGAAAAAAGAGGAGGGYGLGSQTSRGLGCGG 430
Db 542 VGAGNGGNAGTGVGVNGANGNGGSATGALAAVGA-----CAAGDATSGTGGFG--G 594
QY 431 AGAAAAAAGAGGAGGAGGGYGLGSQTSRGLGCGGAGAGAAAAAAGGAG--QGGYG 489
Db 595 AGGSARGLIFALGGAGAA--GGDASTGVGPGPGGTGTASSPFGTATGAGAGAGGAG 652
QY 490 -----GLGSQTSRGGYGPQ-OTSGRGLGCGGAGAGAAAAA 528
Db 653 TSGATGGAGDGVFEGIAVLGLFGGAAGAGGAATGDGATGGAGGFGAGAGIANFLGFS 712
QY 529 AAGGAGGGYGLGSQTSRGLGCGGAGAGAAAAAAGAGGAGGGYGLGSQTSRGG 588
Db 713 VLHG-GAGGAGGTAT-GTGGNGGAGGGGLSPVILGIGGAGGDDGGALGVLG--GNG 768
QY 589 GLGGGAGAA-----AAAAAAGGAGGGYGLGSQTSRGG--GLGGGAGAGAAAAA 642
Db 769 GDDGGGEAVAVGIAVGGAGGAGGAAPTNGGAGGNGGDALGLVGVGNGNAGTGFGAN 828
QY 643 AAGGAGG-----GYGLGSQTS-GPGGYG 667
Db 829 TGGNGGDTTIVVNGMLAPSTLGYGNGGNGVNGGAGGTC 867

Search completed: December 18, 2002, 16:22:16
Job time : 30.4171 secs


```
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 662
FT CONFLICT 672 672
FT CONFLICT 695 747
FT
FT
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match
Best Local Similarity 60.3%; Score 2138; DB 1; Length 747;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

QY 25 QGAGAAAAAAGAGAGAGGGGGLGSGQT--SGRGLGGGAG-AAAAAAGAGGAG 81
D 1 QGAG----AAAAAGAGAGGGGGLGGAGGGGGYGLGGAGGAGGAGAAAAAGAGG 56
QY 82 GYGGLGSGTSGRGLGGGAGAAAAAAGAGAGGGYGLGGSGTSGRGLGGGAG 141
D 57 GYGGLGSGG----AGRGQGAG----AAAAAGAGGGYGLGGSG-AGRGLGGGGA 107
QY 142 GAAAAAAGAGAGGGYGLGGSGTSGPGYGPQQTSGRGLGGGAGAAAAAAG 201
D 108 G----AAAAAGAGGGYGLGNQG-----AGRGQG-----AAAA 142
QY 202 AGAGGGYGLGGSGTSGRGLGGGAGAAAAAAG-----AGAGGGYGLG 252
D 143 AGAGGGYGLGGSG-AGRGLGGGAGAAAAAAGAGGGYGLGGGAGGGYGLG 201
QY 253 SGTSGRGLGGGAGAAAAA-----AAAAAGAGGGYGLG 291
D 202 SQ-AGRGLGGGAGAAAAAAGAGGGGAGGAGAGAAAAAGAGGGYGLG 260
QY 292 QTSRGLGGGAGAAAAAAGAGAGGGYGLGGSGTSGPGYGP-GQTSRGLG 350
D 261 QG-----AGRGEGAG-----AAAAAGAGGGYGLGGG-AGGGYGLGGAGG 311
QY 351 GCGAGAAAA-----AAAAAGAGGGYGLGGSGTSGRGLGGG 392
D 312 GCGAGAAAAAGAGGGGAGGAGGAGAAAAAGAGGGYGLGGSG-AGRGLGGG 370
QY 393 AGAAAAAAGAGAGGGYGLGGSGTSGRGLGGGAGAAAAAAGAGAGGGYGG 452
D 371 AG---AAGAGAGAGGGYGLGGSG-----AGRGQGAG-----AAGAGAGGGYGG 419
QY 453 LGSQTSRGLGGGAGAAAAAAGAGAGGGYGLGGSGT-----SGP 499
D 420 LQNG-AGRGLGGGAG---AAAAAGAGGGYGLGNQAGRGAGGAGGAAAAAGGAG 475
QY 500 GYGP-GQTSRGLGGGAGAAAAAAGAGAGGGYGLGGSGTSGRGLG 552
D 476 GYGGLGSGAGR---GGGAGAAAAAAGAGAGGGYGLGGSG-AGRGLG 531
QY 553 GCGAGAAAA-----AAAAAGAGGGYGLGGSGTSGRGLG 592
D 532 GCGAGAAAAAGAGGGGAGGAGGAGAAAAAAGAGGGYGLGGSG-----AGRG 587
```

```
QY 593 QGAGAAAAAAGAGAGGGYGLGGSGTSGRGLGGGAGAAAAAAGAGAGGGY 652
D 588 QGAG----AAAAAGAGAGGGYGLGGGAG-GRGLGGGAG-----AAGAGAGGGY 636
QY 653 GGLGSGTSGPGYGPQQTSGIRRAPSTS 683
D 637 GGVGASAA-----SAAASRLSSPQAS 659

RESULT 2
YZ08_MYCTU
ID YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTU023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022022; CAAL17745.1; -.
DR TubercuList; RV3508; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 36.2%; Score 1281.5; DB 1; Length 1901;
Best Local Similarity 39.5%; Pred. No. 6.5e-51;
Matches 354; Conservative 28; Mismatches 273; Indels 241; Gaps 39;

QY 7 GOOMGRGSNAGSGGLGGQ-----GAGAAAAAAGAGAGGGYGLGG 52
D 280 GCGGAGGAGGAGGAGGAGGGTGGTGGAGGAGGALLLGGGAGGGGTGGAG 339
QY 53 -----GTSRGLGG-----QGAGAAA-----AAAAAAGG-AGCGGGGLG 90
D 340 GDSVGGVGGTGGGGVAGLGGAGGAGGAGLFGNAGGAGVGGTGGGAGGAGAA 399
QY 91 G-----TSGRGLGGGAGAAAAAAGAGAGGGYGLGGSGTSGR--- 133
```

```
Db 400 GADAPASTGLTGTGTFAGGAGGCGGGNAIAGGINSGGAGGTGGGGAGGSGSADN 459
QY 134 -----GGLGGGQ--AGAAAAAAAAGG-----AGGGYGLGSGQT 168
Db 460 ASGTGADGAGGAGTGTGGVGAAGRAGTGGTGGCGGAGGAGSAGT 519
QY 169 -----SGPGYGPQQTSGRGLGGGAGAAAAAAGAGGQ---- 207
Db 520 DATATGATGTGTFSGAGGAGGAGTGTGGTNGSGGQ--GTGAGGAGGAGGADNPT 578
QY 208 --GGYGLGSGQTSRGLGGQ--AGAAAAAAAAGG--AGGGYGLGSGQT---SGR 259
Db 579 GIGTGTGCKG--GAGGAGGSGSAGGTNGSGAGGTGGGAGGAGGADNPTGI 636
QY 260 GGLGGGAGAAAAAAGAGGAGGCGG--LGSQGTSGRGLGGQ--GAGAAAAA 311
Db 637 GGAGGTGTGGAAGAGGAGGAGTGTGGTGGAVSVNAGIGTGTGGVGGAGGAGAAAA 696
QY 312 AAAAAGGA---GOGYVGLGSG---GTSPPGYGPQQTSGRGLGGQ--- 355
Db 697 GSSATGAGFAGGAGGAGGAGGAGTGTGGTNGSGGAGGAGGAGTGTGGAGGAGDNP 756
QY 356 -----GAAAAAAGGA--GOGYVGLGSG---GTSPPGYGPQQTSGRGLGGQ--- 393
Db 757 AGGAGTGTGGAAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAGGAG 816
QY 394 -----GAAAAAAGGA--GOGYVGLGSG---GTSPPGYGPQQTSGRGLGGQ--- 435
Db 817 FDGGGAGGAGGAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAG 876
QY 436 -----GAAAAAAGGA--GOGYVGLGSG---GTSPPGYGPQQTSGRGLGGQ--- 475
Db 877 NGNAGVGLTAKAGDGAAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAG 936
QY 476 AAAAAGGA---GOGYVGLGSGQTSRGLGGQ--GOGYVGLGSGQTSRGLGGQ-- 531
Db 937 SINAGGAGGAGTGTGGAGGAGTGTGGVGTGAGTGGAGGAGGAGGAGGAGGAG 993
QY 532 GAGGAGGAGGAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAG 580
Db 994 GGGNGRGGGAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAG 1053
QY 581 SGTGSG---RGLGGQ--GAGAAA-----AAAAAGGAGGAGGAGGAGGAGGAG 621
Db 1054 ATGAGGAGGAGGAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAG 1113
QY 622 --TSRGLGGQGA---GAAAAAAGGA---GOGYVGLGSGQTSRGLGGQ 667
Db 1114 NFGGAGGAGGAGGAGGAGGAGTGTGGTGGVGTGAGTGGAGGAGGAGGAGGAG 1169

RESULT 3
FB0H_BOMMO STANDARD; PRT; 5263 AA.
ID FB0H_BOMMO
AC P05790; Q26379; Q17220;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
```

```
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions.";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kinshu X Showa;
RX MEDLINE=89094868; PubMed=3210244;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Biol. 203:917-925(1988).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RX STRAIN=J-139;
RX MEDLINE=99296390; PubMed=10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Biochim. Biophys. Acta 1432:92-103(1999).
CC -I- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -I- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -I- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -I- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-N) INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF226688; AAF76983.1; --
DR EMBL; V00094; CAA3432.1; --
DR EMBL; V00097; CAA3433.1; --
DR EMBL; S74439; AAB31861.1; --
DR EMBL; X13869; CAA32076.1; --
DR EMBL; M35378; AAA27839.1; --
DR EMBL; AB017362; BAA33147.1; --
DR PIR; S01844; S01844.
KW Silk; Signal; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 5263 FIBROIN HEAVY CHAIN.
FT DOMAIN 149 5206 HIGHLY REPETITIVE.
```

```
FT DISULFID 5244 5244 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 5260 5263 C -> V (IN REF. 2).
FT CONFLICT 10 10
SQ SEQUENCE 5263 AA; 391586 MW; 8EE11D3A0A47440E CRC64;

Query Match 33.1%; Score 1174.5; DB 1; Length 5263;
Best Local Similarity 39.6%; Pred. No. 6.3e-46;
Matches 318; Conservative 49; Mismatches 305; Indels 131; Gaps 27;

QY 2 ASMTGGQQMGSGS-MASGRGLGGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGGL 60
DB 4337 ACAGAGYAGAGSGAASGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 4396
QY 61 GQGAGAAAAAAGAGAGGAGG-----GYG-----GLGSGQTSGRGLGGGAGAAAAA 112
DB 4397 SGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4456
QY 113 AAGGAGGGYGGGLGSGQTSGRGLGGGAGAGAAAAAAGAGAGGAGG-GYG-GLGSGQTSG 170
DB 4457 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4516
QY 171 PG---GYGPGQQTSG-----RGGLG-----GOGAGAAAAAAGAGAGGAG 207
DB 4517 ACAGAGAGAGTSSGFGPVANGVYSGYEWYSSRDFCTGSGAGAGSGAGAGAGY 4576
QY 208 G-----GYG-----GLGSGQTSGRGLGGGAGAAAAAAGAGAGGGYGGGLG 253
DB 4577 GAGVAGAGYAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4636
QY 254 QGTSRGLGGGAGAAAAAAGAGAGGAGG-----GYG----- 287
DB 4637 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4696
QY 288 -----GLGSGQTSGRGLGGGAGAAAAAAGAGAGGGYGGGLGSGQTSGPG-GYGP 340
DB 4697 GSGAGAGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4756
QY 341 GQQTSGRGLG-----GOGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGLGGG 391
DB 4757 GYGIGVAGYAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4816
QY 392 GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 443
DB 4817 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4876
QY 444 GAGGGYGGGLGSGQTSGRGLGGGAGAAAAAAGAGAGGAGG-----GYG-GL 491
DB 4877 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4936
QY 492 GSGQTSGPG----GYGPGQQTSGRGLGGGAGAAAAAAGAGAGGGYGGGLGSGQTSG 548
DB 4937 GAGYAGAGAGYAGAGYAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4995
QY 549 RGLGSGGAGAAAAAAGAGAGGAGG-GYG-GLGSGQTSGRGLGGGAGAAAAAAGAG 606
DB 4996 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5055
QY 607 GGAGGQ-----GYG-GLGSGQTSGRGLGGGAGAAAAAAGAGAGGAGG---GYG 653
DB 5056 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5115
QY 654 -GLGSGQTSGPG-GYGPQQTSG 674
DB 5116 AGAGSGAASGAGAGAGAGTSSG 5138

RESULT 4
SPD2_NEPCL
ID SPD2_NEPCL STANDARD: PRT; 627 AA.
AC P46804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Spidroin 2 (Dragline silk fibroin 2) (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.
OX NCBI_TaxId=6915;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=92406876; PubMed=1527052;
RA Hinman M.B., Lewis R.V.;
RT "Isolation of a clone encoding a second dragline silk fibroin.
RL J. Biol. Chem. 267:19320-19324(1992).
CC -!- FUNCTION: Spiders major ampullate silk possesses unique
CC characteristics of strength and elasticity. Fibroin consists of
CC pseudocrystalline regions of antiparallel beta-sheet interspersed
CC with elastic amorphous segments.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- DOMAIN: Highly repetitive protein characterized by regions of
CC polyalanine and glycine-rich repeating units.
CC -!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92913; AAA29381.1; -.
KW Silk; Repeat.
FT DOMAIN 1 530 15 APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 36
FT REPEAT 2 79
FT REPEAT 3 80
FT REPEAT 4 121
FT REPEAT 5 172
FT REPEAT 6 173
FT REPEAT 7 213
FT REPEAT 8 252
FT REPEAT 9 253
FT REPEAT 10 283
FT REPEAT 11 284
FT REPEAT 12 317
FT REPEAT 13 318
FT REPEAT 14 359
FT REPEAT 15 360
FT REPEAT 16 391
FT REPEAT 17 428
FT REPEAT 18 429
FT REPEAT 19 464
FT REPEAT 20 465
FT REPEAT 21 488
FT REPEAT 22 489
FT REPEAT 23 515
FT REPEAT 24 516
FT REPEAT 25 530
SQ SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;

Query Match 32.1%; Score 1138.5; DB 1; Length 627;
Best Local Similarity 47.0%; Pred. No. 7.6e-45;
Matches 325; Conservative 36; Mismatches 190; Indels 141; Gaps 36;

QY 23 GQGGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGLGGGAGAAAAAAGAG-----G 78
DB 2 GEGYGPQ-----QGPGYGP-GQGGPSGPG-----SAAAAAAGAGPGYGP 42
QY 79 AGQGGYGGGLG-SQGTSGRGLGGGAGAAAAAAGAGAGGGYGGGLGSGQ-----GTSR 133
DB 43 PQQQPGGYGPGQPGYGPYGPQGGPSGSAASAAA-GSGQGGPGYGPQGGPGYGP 101
QY 134 GGLGGGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGLGGGAGAAAAAAGAG----- 187
DB 102 CQQGPGSGSAASAAAAGAGAGGGYGGGLGSGQGGPGYGP-PGQQPGYGPQGG 158
QY 188 -GQAGAAAAAAGAGAGGGYGGGLGSGQTSGR-----GLGGGAGAAAAAAGAG 242
DB 159 SGPGSAASAAAAGAGAGGGYGGYGP-GQGGPGYGPQGGPSGSAASAAAAGAG 217
QY 243 AGQGGYGGGLGSGQTSGRGLG-GQ-----GAGAAAAAAGAGAGGGYGGGLGSG 297
```

```

Db 218 QGPGYG-----PGQGGPGYGPQGGGLSGPGSAAAAAAGPGQGGPGYG- GQGGPSGP 272
QY 298 GGLGGGAGAAAAAAGAGAGGCGGCGGGLGSGQTSRGGYGPQQTSGRGGGLGGGAG 357
Db 273 G-----SAAAAAAGAGPGYG-----PGQGGPGYGPQGG-----GPGSAGS 310
QY 358 AAAAAAAGAGGCGGCGGGLGSGQTSRGG-----GLGGGAGAAAAAAGAGAGGAGGAG 412
Db 311 AAAAAAGPGQGGGLGYP-GQGGPGYGPQGGPGYGPQGGPGYGPQGGPGYGPQGGPGY 369
QY 413 GGLGSGQTSRGGGLGGGAGAAAAAAGAGGCGGCGGGLGSGQTSR-----RGGLGGGAG 470
Db 370 GP-GQGGPSGP-----SASAAAAAAGAGPGYGP-GQGGPGYGPQGGPGSGP 417
QY 471 AAAAAAAGAGGCGGCGGGLGSGQTSRGGYGPQQTSGRGGGLGGGAGAAAAAAGAG 530
Db 418 SASAAAAAAGAGPGYG-----PGQGGPGYGPQGGPGYGPQGGPGYGPQGGPGY 463
QY 531 GGAGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGAGGCGGCGGGLGSGQTSRGGGL 590
Db 464 --AGPGYGP-AQGGPSGP-----IAASAAAGPGYGP-----PAQGGP 500
QY 591 GGAGGAGAAAAAAGAGGCGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGAG 647
Db 501 AGYGPSSAASA-----GAGSAGYGP-GSQAASAAASRLASPDSCARVASVNLVSGPT 555
QY 648 QGGYGGGLGSGQTSRGGYGPQQTSGIRIRA 679
Db 556 SSAALSSVISNAVSQIGASNPGLSGCDVLQA 587

RESULT 5
WA22_MYCTU
ID WA22_MYCTU STANDARD; PRT; 914 AA.
AC 006794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE WAG22 antigen precursor.
GN WAG22 OR RV1759C OR WT1807 OR WTCY28.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekalia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

```

```

CC FRAMESHIFT IN POSITION 85.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95890; CAB09322.1; -.
DR EMBL; AE007040; -. NOT_ANNOTATED_CDS.
DR TIGR; MT1807; -.
DR TubercuList; Rv1759c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
KW Antigen; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAG22 ANTIGEN.
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8E6AC8 CRC64;

Query Match 31.3%; Score 1110.5; DB 1; Length 914;
Best Local Similarity 38.2%; Pred. No. 1.6e-43;
Matches 309; Conservative 31; Mismatches 307; Indels 161; Gaps 33;

QY 5 TGGQMGGRGSMASRGGLGGGAGAAAAAAGAGAG-----QGGYGGGL-GSQGTSRGG 59
Db 110 TGRPLGNG--ANGAPGTGANGDAGWLIGNGAGGSCAKGANGGAGGPGGAGLFGNGG 167
QY 60 LGGOGAGAAAAAAGAG-----GQGGYGGGLGSGQTS-----GRGLGGGAGAAAAA 109
Db 168 AGAGGTATANNIGTGAGGAGGGSAMLFAGAGGAGGAGGAATSLVGGIGTGTGTGNAGMLA 227
QY 110 AAAAAGAG-----QGGYGGGLGSGQ-----TSRGGGLGGQ-----ACAAAAAAGG 154
Db 228 GAAGAGGAGGFSFTAGGAGGAGGAGGLFTTGGYGGAGGQGTGGAGGAGGAGLFGAGG 287
QY 155 -AGQGGYGGGLGSGQTSRGGYGPQQTSGRGGGLGGGAGAAAAAAGAG-----204
Db 288 MGGAGGFGDHTLGTGGAGGDDGGGLFGAGGDDGAGGSGLTTCGAAGNGNAGTSLIGA 347
QY 205 -AGQGGYGGGLGSGQ-----SGRGGGLGGGAGAAAAAAGAGAG-----OGYGGLSG-- 253
Db 348 AGGAGGTGGAG--GTFVGGKGGAGGAGGAGNAGMLFGSGGGGGTGGFFMAGGQGGVGSA 405
QY 254 ---QGTSGRGLGGQ--ACAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 306
Db 406 GMLSGSGSGGAGGSGGPGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 465
QY 307 A-----AAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 327
Db 466 AVLIGNGEGGIGALAGKSGFGGGLLILCADGYNAPESTSPWHNLOQDILSFINEPTEA 525
QY 328 -----GSGTSGPG-GYGPQQTSGRGGGLGGGAGAAAAAAGAG-----GQG 372
Db 526 LTRPLTNGDSGTPGTDDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 584
QY 373 GYGLGSGQTSRGGGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 432
Db 585 GAGGAGGAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 634
QY 433 AAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 487
Db 635 AGLLIGAAGLGGCGGAGTAFTAVTTGGAGGTCGAAGLFGANGGAGGAGGAGGAGGAGGAGG 694
QY 488 YGGLGSG--GTSGPGYGPQQTSGRGGGLGGGAGAAAAAAGAGAG-----QGGYGGLSQ 544
Db 695 AGGLYHGHTGGGNGGSGTGTAGGTGGAGGPG--GLYAGGSGGAGGAGGAGGAGGAGGAGG 753
QY 545 GTS-----GRGGLGQ-----GAGAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 591
Db 754 AGSITLNASGAGGSGGSLSGRAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 813

```

```

QY 592 GGG-----ACAAAAAAGAGAGAGGGGG-IGSOCTSGRGLGGGAGAAAAA 641
Db 814 GAGQITGLRSGRGGAGGAGADTGAGNGGAGGAGGAGGAGGATSGKGGAG 873
QY 642 AAAGAGAGGGGGLGSGQ--GTSPPGGYG 667
Db 874 GNAVVGNGGNGAGKAGGTAGAGGAG 901

RESULT 6
YQ34_MYCTU
ID YQ34_MYCTU STANDARD; PRT; 778 AA.
AC P71933;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv2634c.
GN RV2634C OR MT2712 OR MTCY441.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fieischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; Z80225; CAB02341.1; -
DR EMBL; AF007103; AAK47036.1; ALT_INIT.
DR TIGR; MT2712; -.
DR TubercuList; RV2634c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 29.6%; Score 1049.5; DB 1; Length 778;

```

```

Best Local Similarity 41.3%; Pred. No. 7.5e-41;
Matches 291; Conservative 21; Mismatches 294; Indels 99; Gaps 34;

QY 7 GQMGGRGSMASRGGLGGGAGAGAAAAAAGAGAGAGGGGGLGSGQ-----GTSGR 57
Db 129 GAPGPGGGLLLGNGNGSGAPGPGGAGGDAGLIGNGTGTGKGGDLVGSAGAGVGGR 188
QY 58 GG--LGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGRGLGGGAGAAAAAAG 115
Db 189 GCVLLNGGTGCGAGGAATLVG-GTGGVG--GATGLISGGFG--GAGAAAGVTTGG 243
QY 116 GAGQGGYGGGLGSOCTSGRGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGPGVG 175
Db 244 VGGSGVGCV--FGNGGFGGAGLGAAGCVGAASYFTGTGGGGVGGDAPG--GDGGAG 299
QY 176 PQQQTSGRGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGRGLGGGAGAAAAA 235
Db 300 P--LLINGVGGLG-GAGAA-----GNGGAGGM-----LLGGGAGGCGGPAVAGV 344
QY 236 AAAAAAGGAGGQ-----GYGGLGSOCTSGRGLGGGAGAGAAAAAAG-----GAGQ 283
Db 345 LGGMPGAGGNGGNANWFGSSGAGGGGTGCTAGTNGVNPCSIANTGANGTDSNGNQ 404
QY 284 GGYGGLGSOCTSGR--GGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGPGYGPQ 342
Db 405 GGNPGPGPAGGYGEAGVGCGQ-----GLGESLDGNDGTGKGGAG--CTAGTDG-GAG- 455
QY 343 QTSRGGGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGRGLGGGAGAAAAAAG 402
Db 456 ---GAGGAGGTCTGCSAGCVATGEGGDDATGVDG-GVGAGGKGGGHTGVDAPG 511
QY 403 AAGAGQGGYGGGLGSOCTSGRGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSG- 460
Db 512 GDGGIGGDNALGAAG--GNGTGGAGNGGGRGGLGNGGAGGAGGTGTTGGGGAAGF 569
QY 461 RGLGGGAGAGAAAAAAGAGAGGGGGLGSGOCTSGPGYGPQQQTSGRGLGGGAG 520
Db 570 AGGVGAGGEGLTDTAGTAEGGT--GGLGLG--GVGTTGGMG-----GSGVGNGGA 619
QY 521 AAAAAAAGAGAGAGGGGGLGSGOCTSGRGLGGGAGAGAAAAAAGAGAGGGG 578
Db 620 AGSLGLGGGGAGGVTGGTGG--GIGAGNGGAGGAGGTTGGGATIGGGGGTGGVG 677
QY 579 LGSOCTSGRGLGGQ--GAGAAA-----AAAAAAGAGAGGGGGLGSGOCTSGRGLGGQ 631
Db 678 AG--GTGTTGAGGTTGSGGAGGLIGWAGAGGTGAGTGGGGGLGGQ--GNGGNGT 733
QY 632 GAGAAAAAAGAGAGGGGAGG-----LGSOCTSGPGYGPQQ 671
Db 734 GATGGGGDFALGNGGAGGAGGSPGSGSIQGNMGP-----PGTQ 774

RESULT 7
Y278_MYCTU
ID Y278_MYCTU STANDARD; PRT; 957 AA.
AC P56877;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv0278c precursor.
GN Rv0278C OR MT0291 OR MTV035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA

```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AL021958: CAA17514.1; -;
DR EMBL: AE006968: AAK45011.1; -;
DR TIGR: MT0772.5; -;
DR InterPro: RV0747; -;
DR Pfam: PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 801 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV0747.
FT CONFLICT 188 G -> S (IN REF. 2).
FT CONFLICT 225 R -> G (IN REF. 2).
FT CONFLICT 227 R -> G (IN REF. 2).
FT CONFLICT 295 K -> R (IN REF. 2).
FT CONFLICT 300 S -> G (IN REF. 2).
FT CONFLICT 338 T -> I (IN REF. 2).
FT CONFLICT 377 A -> P (IN REF. 2).
FT CONFLICT 577 T -> A (IN REF. 2).
FT CONFLICT 580 MISSING (IN REF. 2).
SQ SEQUENCE 801 AA; 65407 MW; EA54C9BF45A00F41 CRC64;

Query Match 27.9%; Score 988.5; DB 1; Length 801;
Best Local Similarity 38.3%; Pred. No. 3.8e-38;
Matches 277; Conservative 35; Mismatches 305; Indels 107; Gaps 28;
QY 11 GRGSMASGRGLGGOGAGAAAAAAGGACGGYGGGLGSGQTSRGLGGGAGAAAA 70
DB 111 GRPLNGANGAPGTGANGPGGWLINGGAGGSGAPGAGAGNGGAGGLFGSG-GAGGA 169
QY 71 AAAAAAGGACGGYGGGLGSG--GTSGRGLGGOGAGAAAAAAGGACGGYGGGLGSG 128
DB 170 STDVAGGAGGAGGAGNAGMLFGAAGVGGVGFNSG-GATGAGAGGAGG-GLFGAGR 227
QY 129 GTSGRGLGGOGAGAAAAAAGGACGGYGGGLG-SQTSRGGPGY-GPGQQTSGRGL 186
DB 228 GSGGSGNLTGGAGGAGNAGTLATDGGAGGTGASRGGFGGAGGAGGAGGAGGAGG 287
QY 187 GGOGAGAAAAAAGGACGGYGGGLGSGQTSRGL--CGGAGAAAAAAGGA 243
DB 288 GGAG-GISKVSDSAAGGAG-CAPGLINGNGGNGGASTGGGDPGCGAGGTGLING 345
QY 244 GGGYGGGLGSGQTSRGLGGOGA----- 267
DB 346 GNGSGGTGA--TLCKAGIGGTGGVLLGLDGFAPASTPLHTLOQDVINWYNDPFTLT 403
QY 268 -----GAAAAAAGGAGGAG--GYGGLGSGQTSRGLGGOGAGAAAAAAGG 318
DB 404 GRPLICNGANGTPTGTGADGAGGWLFGNGNGGOGTI--GGVNG-GAGGAGGAGILFT 460
QY 319 AGCGYGGGLGSGQTSRGGYGPQQTSGRGLGGOGA-----GAAAAAAGGAGG 371
DB 461 GTGGSGGPGATGLGIGGAGGAALLFGSGAGSGGAGVGGNGGAGNGAGALLGAAGA 520
QY 372 GGYGGLGSGQTSRGLGGOGAGAAAAAAGG----AGGGYGGGLGSG--GTSRGL 426
DB 521 GGAGGAGV--GNGAGNG--GLFANGGAGCGPGFGSPAGAGGLGAGNGGGLFCAGGT 577
QY 427 GGOGAGAAAAAAGGACGGYGGGLGSGQTSRGLGGOGAGAAAAAAGGAGGAG 486
DB 578 GGAGGSGTLAGGAGGAGNG-----GLFGAGGTGGAGSHSTAAGVSGAGGA--G 625
QY 487 GYGGGLGSGQTSRPGYGPQQTSGRGLGGOGAGAAAAAAGGACGGYGGGLGSGQ 546
DB 626 DGAGLLSLGSGAGGAGGSGSLTAAGVVGIG-----GAGGLLFGSGGAGSGGFSNSGN 680
QY 547 SGRGLGSGQ-----GAGAAAAAAGGAGGAGYGGGLGSGQTSRGLGGOGAGAAAAA 601
DB 681 GGAGGAGGADAGLLVSGGAGGAGASAT-CAATGGGAG--GKSAFGLGGDGGAGGATG 737

QY 602 AAAAAAGGACGGYGGGLGSGQTSRGLGGOGAGAA-AAAAAAGGACGGYGG--LGSQ 658
DB 738 LSGAFHIGKGGVG--GSNAVLIENGNGNGNGSGNAGKSGGAPGSPGAGGAGLLG 795
QY 659 GYSG 662
DB 796 GLNG 799

RESULT 9

ELS_MOUSE
ID ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
GN ELatin precursor (Tropoelastin).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to mouse chromosome 5 in a region of linkage conservation with human chromosomes 7.,"
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U08210; AAA80155.1; -;
DR MGD: MGI:95317; Ein.
DR InterPro: IPR003979; tropoelastin.
DR PRINTS: PR01500; TROPOELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
SQ SEQUENCE 860 AA; 71955 MW; 0C0BESAAE1EDD7F1 CRC64;

Query Match 27.5%; Score 976; DB 1; Length 860;
Best Local Similarity 39.3%; Pred. No. 1.4e-37;
Matches 323; Conservative 36; Mismatches 249; Indels 214; Gaps 40;
QY 13 GSMSASGRGLGGQG-----AGAAAAAAGGACGGYGG 48
DB 78 GTFGAGPGLGGAGPGAGLGPAGTFFPGAGALVPGGAAGAAAYKAAKAGAGLGGVGG 137
QY 49 LGSQTSRGLL-GGOGAGAAAAAAGGACGGYGGGLGSGQTSRGLGGGAGGAAA 107
DB 138 V--PGGVGGVPGVGVG-----GVPGGVGGVGP-----GVGGIGGIGLGVSTCA 184
QY 108 AAAAAAGG-GAG--GGYGGGLGSGQTSRGLGGOGA-----GAAAAAAGG 151
DB 185 VYPVGAGIGAGGKPKVPGVGLPGVPGVLPGTGARFPFGVGLPVGPTGTGAKAPG 244
QY 152 AGGAGG--GYGGLGSG-----GTSGPGCYG-----PGQQTSGRGLGGO----- 189


```

[3]
RP SEQUENCE FROM N.A.
RX MEDLINE-98318805;
RA Beckwith R., Arcidiacono S., Stote R.;
RT "Evolution of repetitive proteins: spider silks from Nephila clavipes
  (Tetragnathidae) and Araneus bicentenarius (Araneidae).";
RL Insect Biochem. Mol. Biol. 28:121-130(1998).
[4]
RP SEQUENCE FROM N.A.
RA Beckwith R.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20329; AAC38957.1; -.
DR HSSP; P10969; 1WGT.
FT NON_TER 1
SQ SEQUENCE 544 AA; 44107 MW; CC611B3551945615 CRC64;

Query Match 48.8%; Score 1729.5; DB 5; Length 544;
Best Local Similarity 61.3%; Pred. No. 1.4e-86;
Matches 409; Conservative 12; Mismatches 33; Indels 213; Gaps 28;

QY 18 GRGGLGGGAG-AAAAAAAAAGGAGGGGGGGLGSGTSGRGLGGGAGAGAAAAA 76
DB 1 GYGLGGGGAGGAGAAAAAGAGGGGGGLGSG-AGRGGGAG-AAAAA 52

QY 77 GGAGGGGGLGSGTSGRGLGGGAGAAAAAGAGGGGGGGLGSGTSGRGL 136
DB 53 GGAGGGGGLGSG-AGRGGLGGGAG-AAAAAGGVG-GLL 92

QY 137 GGAGAGAAAAAGAGGGGGGGLGSGTSGPQGTSGRGLGGGAGAGAAA 196
DB 93 GGAGAGGAGAAAAAGAGGGGGGLGSG-AGRGSGGGAG-135

QY 197 AAAAAAGAGGGGGLGSGTSGRGLGGGAGAAAAAGAGGGGGGGLGSGT 256
DB 136 AAAAAAGAGGGGGLGSG-AGRGGLGGGAG-AAAAAGAGGGGGGGLGSG-189

QY 257 SGRGLGGGAGAAAAAGAGGGGGGGLGSGTSGRGLGGGAGAAAAA 316
DB 190 -----GAGGGGGGLGSG-AGRGGLGGGAG-AAAAA 220

QY 317 GGAGGGGGLGSGTSGPQGTSGRGLGGGAGAAAAAGAGGGGGG 376
DB 221 GGAGG-----GGLGGGAG-AAAAAGGAGGGGG 250

QY 377 LGSQTSRGLGGGAGAAAAAGAGGGGGGGLGSGTSGRGLGGGAGAAA 436
DB 251 LGSQ-----AGRGGGAG-AAAAAGGAGGGGG-280

QY 437 AAAAAAGGAGGGGGGGLGSGTSGRGLGGGAGAAAAAGAGGGGGGGLGSGT 496
DB 281 -----GAGGGGGGGLGSG-AGRGGLGGGAG-AAAAAGGAGG-320

QY 497 SGPGGYGPQGTSGRGLGGGAGAAAAAGAGGGGGGGLGSGTSGRGLGGG 556
DB 321 -----GGLGGGAG-AAAAAGGAGGGGGG-AGG 351

QY 557 AAAAAAGAGGGGGGGLGSGTSGRGLGGGAGAAAAAGAGGGGGGG 616
DB 352 AAAAAAGAGGGGGGGGGGGLGSG-AGRGGGAG-AAAAAGGAGGGGG 403

QY 617 LGSQTSRGLGGGAGAAAAAGAGGGGGGGLGSGTSGPQGTSGIR 676
DB 404 LGGGCV-GRGLGGGAGAAAV-----GAGGGGGGGGAG-AAAAA 448

QY 677 IRAPSTS 683
DB 449 LSSPQAS 455

RESULT 4
Q9BIU3 PRELIMINARY; PRT; 691 AA.
AC Q9BIU3;
```

```

01-JUN-2001 (TrEMBLrel. 17, Created)
01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Fibroin 2 (Fragment)
OS Dolomedes tenebrosus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Lycosoidea; Pisauridae; Dolomedes.
OX NCBI_TaxID=156846;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
  Fibroin Sequences.";
RL Science 291:2603-2605(2001).
DR EMBL; AF350270; AAK30599.1; -.
DR InterPro; IPR00104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01228; EGSHELL.
FT NON_TER 1
SQ SEQUENCE 691 AA; 56176 MW; CB46BF9F1991F33E CRC64;

Query Match 45.0%; Score 1594.5; DB 5; Length 691;
Best Local Similarity 56.6%; Pred. No. 3.2e-79;
Matches 406; Conservative 39; Mismatches 89; Indels 183; Gaps 41;

QY 24 GOGAGAAAAAGAGGGGGGLG-----SQTSGRGLGGGAGAAAAA 71
DB 2 GOGSGAGAAAAAGAGGGGGGPGYASLYLSTTTTSSOGAGGGGGGGSGTGSNA 61

QY 72 AAAAAAGAGGGGGGLGSGTSGRGLGG-----GAGAAAAAAMAG-GAGGGGGGLG 126
DB 62 AAAAAAGAGGGGGGGG-AG-AGGGGGGGGGGAGAAAAAGAGGGGGGG-116

QY 127 SQTSGRGLG--GOGAGAAAAAAG-GAGGGGGGLGSGTSGPQGTSGR 183
DB 117 -----GOGGLGGGGAGAAAAAGAGAGGGGG-AGGGGG-157

QY 184 GGLGGGAGAAAAAAG-GAGGGGGGLGSGTSGRGLG--GOGAGAAAAA 240
DB 158 --GGAGAAAAAAGAGGGGGGG-AGGGGGGGGGAGAGAGAGAGAGAGAG 206

QY 241 G-GAGGGGGGLGSGTSGRGLGGGAGAAAAAAG-GAGGGGGGLGSGTSGR 298
DB 207 GAGAGGGY-----GOGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 253

QY 299 GLG-GOGAGAAAAAAG-GAGGGGGGLGSGTSGPQGTSGRGLGGGAG 356
DB 254 GYGGAGAGAAAAAGAGAGGGGGGG-AGGGGG-----GOGAG 295

QY 357 AAAAAAGAGGGGGGGGLGSGTSGRGLGGGAGAAAAAAG-GAGGGGGGL 415
DB 296 AGAAAAAGAGAGGGGG-AGGGGG-AGAAAAAAGAGGGGGGGGGGG 348

QY 416 GSQTSRGLG--GOGAGAAAAAAG-GAGGGGGGLGSGTSGRGLG--GOGAGAA 472
DB 349 G-----GLGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398

QY 473 AAAAAAGAGGGGGGLGSGTSGPQGTSGRGLGGGAGAAAAAAGG 532
DB 399 AGAGSGGGAGGGGG-AGGGGG-----GOGAGAGAGAGAGAGAG 434

QY 533 AGGGGGGLG-----SQTSGRGLGGGAG-AAAAAAGAGGGGGGG-577
DB 435 AGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 492

QY 578 -GLGSGTSGRGLG--GOGAGAAAAAAG-GAGGGGGGLGSGTSGRGLG--GQ 631
DB 493 AGAGAGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 544

QY 632 GAGAAAAAAG-GAGGGGGGLGSGTSGPQGTSGPQGTSGPQGTSGPQGT 682
DB 632 GAGAAAAAAG-GAGGGGGGLGSGTSGPQGTSGPQGTSGPQGTSGPQGT 682
```

```
Db 545 GAGAGASAAASAGSGGGG--GQGGYGGTGTGAGAAASSAGVAVTVGNT 593
RESULT 5
Q9BIU7
ID Q9BIU7 PRELIMINARY; PRT; 648 AA.
AC Q9BIU7
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Major ampullate spidroin 1 (Fragment).
GN MASp1.
OS Argiope trifasciata.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Argiope.
OX NCBI_TaxID=156845;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences."
RL Science 291:2603-2605(2001).
DR EMBL: AF350266; AAK30595.1; -.
DR HSP: P10969; 1WGT.
FT NON_TER 1
SQ SEQUENCE 648 AA; 52887 MW; ACDB67067B2302CB CRC64;

Query Match 44.3%; Score 1568; DB 5; Length 648;
Best Local Similarity 58.5%; Pred. No. 8.1e-78;
Matches 388; Conservative 35; Mismatches 128; Indels 112; Gaps 30;

QY 29 AAAAAAAAAAGGAGGGGGLGSGTSGRGLGGGAGAGAAAAAAGGAGGGYGLG 88
Db 1 AAAAAAAAAAGGAGGGGGLGSG--AGGGYGGGAAAAAASCA-----46
QY 89 SGTSGRGLGGGAGAGAAAAAAGAGAGGGYGLGSGTSGRGLGGGAGAGAAAA 148
Db 47 --GSAQRGLG-----AGGAGQ-GYG-----AGSGGGGAGGGGAAAT 82
QY 149 AAAAGG-AGGGYGGGLGSGTSGPGYGPQOTSGRGLGGGAGAGAAAAAAGGAG 207
Db 83 AAAAGGGGGYGGGLGSG-SSGGY-----GAGGAGAAAAAAGGAGGAG 128
QY 208 GY--GGLGSGTSGRGLGGGAGAGAAAAAAGG-AGGGYGGGLGSGTSGRGLGG 264
Db 129 EGLGAGGAGGGYAGLGGGGAGAGGAGAAAAAAGGGGGGGYGGGLGSG-AGGGY-G 186
QY 265 QGAGAAAAAAGGAGGGGGGLGSG-AG-TSGRGLGGGAGAGAAAAAAGG-AGG 321
Db 187 QGGAAGAAAAAGGAGGGGGLGAGAGAGGGYAGSGGGGAGGGGAGAAAAAAGGG 246
QY 322 GYGGLGSGTSGPGYGPQOTSGRGLGGGAGAGAAAAAAGGAGGGY--GGLGS 379
Db 247 GYGGLGSG-AGGGY-----GAGGAGAAAAAAGGAGGGGAGGAGGAG 292
QY 380 QTSRGLGGGAGAGAAAAAAGG-AGGGYGGGLGSGTSGRGLGGGAGAGAAAA 438
Db 293 EYGAAGGGGAGGAGAGAAAAAAGGGGGGGYGGGLGSG-AGGGY-GGGAGAAAA 350
QY 439 AAAAGGAGGGY--GGLGSGTSGRGLGGGAGAGAAAAAAGG-AGGGYGGGLGSG 495
Db 351 ASGAGARRGGLGAGAGGGYAGLGGGGAGAGGAGGAGGAGGAGGGGGYGGGLG 410
QY 496 TSGPGYGPQOTSGRGLGGGAGAGAAAAAAGGAGG-AGGGYGGGLGSGTSGR 553
Db 411 --SGGGY-----GAGGAGAAAAAAGGAGGGGAGGAGGGYGGGLGSG 456
QY 554 GAGAGAAAAAAGG-AGGGYGGGLGSGTSGRGLGGGAGAGAAAAAAGGAGGAG 612
Db 457 GAGGAGAAAAAAGGAGGGGGGGYGGGLGSG-----AGGGYGGGAGAAAA 512
QY 613 GYGGLGSGTSGRGLG-----AGAGAAAAAAGGAGGGYGGGLGSGTSGPG 665
```

```
Db 513 GTGGLGSG-AGGGYGGGAGAFSGQGGGAASVATASAAASRLSPGAASRVSAVTSLVSS 571
QY 666 YGP 668
Db 572 GGP 574

RESULT 6
Q9BIU4
ID Q9BIU4 PRELIMINARY; PRT; 854 AA.
AC Q9BIU4
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Fibroin 1 (Fragment).
OS Dolomedes tenebrosus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Lycosoidea; Pisauridae; Dolomedes.
OX NCBI_TaxID=156846;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences."
RL Science 291:2603-2605(2001).
DR EMBL: AF350269; AAK30598.1; -.
DR InterPro: IPR000217; Tubulin.
DR PROSITE: P500227; TUBULIN; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 854
SQ SEQUENCE 854 AA; 71881 MW; 5C52F1B0AE31C6C2 CRC64;

Query Match 43.5%; Score 1539.5; DB 5; Length 854;
Best Local Similarity 50.7%; Pred. No. 3.5e-76;
Matches 412; Conservative 46; Mismatches 145; Indels 209; Gaps 47;

QY 7 GQGMGRGSMASGRGLGGGAGAGAAAAAAGG-AGAGGGYGGGLGSGTSGRGLG--GQ 63
Db 34 GSGGGYDQSSGGY-GGAGAGATATTAAGAGSGGGY-----GQGLGGY 84
QY 64 GAGAAAAAAGG-AGGGY-----GLGSGTSGRGG 97
Db 85 GAGAGAAAAAAGGAGSGGGYDQGGGLGGYGGGAGSATAPAGSGFGGCGFGRNGG 144
QY 98 LG--GAGAG--AAAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAGAAAAAAG 152
Db 145 KGAYGGSAGAGVGAATAAAGAGGGGGY-----DQGLGGYGGGAGAGAGAA 196
QY 153 GGAG--QGGY--GGLGSGT-----SGPGYGPQOTSGRGLGGG 190
Db 197 GGGDYEGGGYGGGLGSGFGGAGAGAAAAAAGGAGSGRGGY-----DQGLGGY 250
QY 191 AGAAAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAGAAAAAAGG-AGGGY- 248
Db 251 QGAGAGAAAAAGGGGGYGGYDQ--GRGY--GQSGAGSATAAAGAGGFGGGY 307
QY 249 --GGLGSGTSGRGLG--GGAGAAAAAAGG-AGGGYGGGLGSGTSGRGLGG 304
Db 308 QGGYGNQ-----GGLGSGGAGAGAAAAAAGGAGSGRGGY-----DQGLGGY 354
QY 305 AGAAAAAAGGAG--GQGGY--GGLGSGT-----SGPGYGPQ 342
Db 355 QGAGAGAAAAAGGGYGGYGGGLGSGFGGAGAGAAAAAAGGAGSGRGGYGGG 414
QY 343 QTSRGLG--GGAGAGAAAAAAGG-AGGGY----- 375
Db 415 Y-GNGLGSGFGGAGAGAAAAAAGGAGSGRGGYDQGLGGYGGAGAGAAAAAGG 473
QY 376 --GLGSGTSGRGLG--GGAGAAAAAAGG-AGGGY--GGLGSGTSGRGLG-- 427
Db 474 DGYGGGGYGGGLGSGFGGAGAGAAAAAAGGAGSGRGGYGGYGNQ-----GGLGSF 528
```

```

QY 428 GCGAGAAAAAAGAG-GAGGGGGLGSGTSGRGLGGGAGAGAAAAAAGGA--- 483
DB 529 GCGAGAAAAAAGAGAGSGRGY-----DQGLGGYGAGAGAAAGGDDGY 580
QY 484 GGGG---GLGSQ-----TSGPGGYPGQOTSG-RGGLGGGAGAAAAAAG 530
DB 581 GGGGYNQRGVSYCGAGAGAAATSAAGGAGSGRGYEQGLGGYGAGAGAASTAA 640
QY 531 GGA---GCGGGGLGSGTSGRGLGGGAGAGAAAAAAGS-ACGGYGGGLSGTSG 586
DB 641 GGGDYGCGGGYNGCGRGSYG---GSGAGAGAAVAAAGAVSGGGYDGG-----G 691
QY 587 RGLGGGAGAGAAAAAAG-GAGGGYGGGLGSGTSGRGLGGGAGAGAAAAAAG 645
DB 692 QGGY-CGGSAGAGAAVAAAGGTAGGGYGGSGQAGYGGGAG---GFRAAATTAAGAG 746
QY 646 GAGGGYGGGLGSGTSGPGGYPGQOTSGIRI 677
DB 747 GAG-GGCGGGYGGGGYGGTGGAGGASSAGLSV 777

RESULT 7
Q9BIT7 PRELIMINARY; PRT; 1953 AA.
AC Q9BIT7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major amputate spidroin 2-like protein (Fragment).
OS Nephila madagascariensis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=115969;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences.";
RL Science 291:2603-2605(2001).
FR EMBL; AF350276; AAK30605.1; -.
FT NON_TER 1
FT NON_TER 1953
FT SEQUENCE 1953 AA; 159383 MW; C82B4DACD043C3BD CRC64;

Query Match 41.0%; Score 1452.5; DB 5; Length 1953;
Best Local Similarity 39.6%; Pred. No. 3.2e-71;
Matches 445; Conservative 21; Mismatches 184; Indels 473; Gaps 50;

QY 2 ASMTGGQQHGRGSMASRGGLGGGAGAAAA-----AAAAAGGAGG 43
DB 852 AAASAGR--GPGGYGPGQGGPGSGSAAAAAAGRGPGGYPGQGGPGGAAAAAAGRG 909
QY 44 GGYG-----GLGSQGTSGR-----GLGGGAGAGAAA----- 70
DB 910 GGYGPGQGGPGGPGGSGSAAAAAAGRGPGGYPGQGGPGGAAAAAAGPGGYPGQGG 969
QY 71 --AAAAAGAGGGYGGGLGSGTSGR-----GLGGGAGAGAAAAAAG 115
DB 970 PGAAAAAAGSGPGGYGPG-GQGGPGGAAAAAAGRGPGGYPGQGGPGGAAAAAAG- 1027
QY 116 GAGGGYG---GLGSQGTSGR-----GLGGGAGAGAAA----- 145
DB 1028 GRGPGGYGPGQGGPGGSGSAAAAAAGRGPGGYPGQGGPGGAAAAAAGPGGYG 1087
QY 146 -----AAAAAAGGAGGGYG---GLGSQGTSG-----GPGGYGPGQOTSGRGLGG 188
DB 1088 GQGGPGAAAAAAGRGPGGYPGQGGPGGAAAAAAGRGPGGYPGQGGPGGQGGPGG 1147
QY 189 QGAGAAAA-----AAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAGAAA- 234
DB 1148 PGAAAAAAGPGGYGPGQGGTGAAGAAAAAGSGAGGYG-----PGQGGPGGCAAAAAAG 1200

```

```

QY 235 -----AAAAAGCAGGGYG-----GLGSQGT----- 256
DB 1201 PGYGPQGGPGAAAAAAGSGPGGYGPGQGGPGSSAAAAAAGPRYGPQGGPGAAAA 1260
QY 257 ----SGRGL--CGGAGAGAAAAAAGCAGGGYGGGLGSGTSGR-----G 299
DB 1261 AAAGSGPGGYGPGQGGPGGAAAAAAGAGRGPGGYGPG-GQGGPGGCAAAAAAGPGGYG 1319
QY 300 LGGQAGAGAAA-----AAAAAGGAGGGYG-----GLGSQ----- 331
DB 1320 PGQGGPGCAAAAAAGSGPGGYGPGQGGPGGAAAAAAGRGPGGYGCGQGGPGGPGAAA 1379
QY 332 -TSGPGGYGPGQ-----TSGRGL--GGGAGAGAAAAAAGCAGGGYG--- 375
DB 1380 AAAGPGGYGPGQGGPGGAAAAAAGSGPGGYGPGQGGPGRSGAAAAAAGRGPGGYGPG 1439
QY 376 -----GLGSQ-----TSGR-----GLGGCAGAGAAA--- 398
DB 1440 QGPGPGCAAAAAAGPGGYGPGQGGPGGAAAAAAGRGPGGYGPGQGGPGGCAAAAAAGRG 1499
QY 399 -----AAAAAGCAGGGYGGGLGSGTSGR-----G 425
DB 1500 PGYGPQGGPGGPGGAAAAAAGTGPGYGPG-GQGGPGGCAAAAAAGRGPGGYGPGQGG 1558
QY 426 LGGQAGAGAAA-----AAAAAGGAGGGYGGGLGSGTSGR----- 462
DB 1559 PGGPGAAAAAAGPGGYGPGQGGPGGAAAAAAGSGPGGYGPG-GQGGPGGCAAAAAAGRG 1617
QY 463 -----GLGGGAGAGAAA-----AAAAAGGAGGGYG---GLG 492
DB 1618 GGYGPGQGGPGGCAAAAAAAGRGPGGYGPGQGGPGGSGAAAAAAGRGPGGYGPGQGG 1677
QY 493 SQ-----TSGPGGYGPGQ-----TSGR-----GLGGGAGAGAAA--- 525
DB 1678 GPCAAAAAAGPGGYGPGQGGPGGAAAAAAGRGPGGYGPGQGGPGGCAAAAAAAGRGPG 1737
QY 526 -----AAAAAGCAGGGYGGGLGSGTSGR-----GLGGCAGAGAAAA 566
DB 1738 GYGPQGGPGGSGAAAAAAGRGPGGYGPG-GQGGPGGCAAAAAAAGPGGYGPGQGGPGAAA 1796
QY 567 AAGGAGGGYGGGLGSGTSGR-----GLGGCAGAGAAA----- 600
DB 1797 AAGRGPGGYGPG-GQGGPGGSGAAAAAAGRGPGGYGPGQGGPGGCAAAAAAAGPGGYGPG 1855
QY 601 -----AAAAAGCAGGGYGGGLGSGTSGRGLGGGAGAGAAA-----AAAA 642
DB 1856 QGCTGAAAAAAGSGAGGYG-----PGQGGPGGCAAAAAAAGPGGYGPGQGGPGGAAA 1908
QY 643 AAGGAGGGYG---GLGSQGTSG-----GPGGYGPGQGTSG 674
DB 1909 AAAGSGPGGYGPGQGGPGGPGGAAAAAAGRGPGGYGPGQGGPG 1951

RESULT 8
OL7434.
ID OL7434 PRELIMINARY; PRT; 988 AA.
AC OL7434;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Minor ampullate silk protein Mispl (Fragment).
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
RT and highly conserved non-silk-like 'spacer regions'.";
RL Protein Sci. 7:667-672(1998).

```



```
Query Match      35.9%; Score 1273; DB 16; Length 1217;
Best Local Similarity 39.0%; Pred. No. 1.1e-61;
Matches 336; Conservative 37; Mismatches 279; Indels 210; Gaps 35;

QY 6 GGOQMGRGSMASGRGLGSGQ-----GAGAAAAA-----AA 35
DB 214 GGGAGGAGGAGGTGAGAGRAELLFGAGGAGAGTGGPCATGCTGGHGVGGDGLWLP 273

QY 36 AAAGGA-GGGYGGGLGSGQ--TSRGGGLGQ-GAGAAAAAAGGAGCGGYYGGLGSG- 90
DB 274 GAGAGGAGGAGGAGSDGALGTGTTGTTGAGGAGGAGGALLLLGAGGQGLGAGGQ 333

QY 91 -----GTSRGGGLGQ-----GAGAAAAAAGGAGCGGYYGGLGSG 129
DB 334 GAGGAGGAGADNPTGIGTGGDGTGGSAGGAGGAGGAGGAGGAGGAGGAGGAGG 393

QY 130 TSGRGLGSGQ-----GAAAAAAGGAG-----GOGYVG 161
DB 394 --GAGGAGGAGADADPGATGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 451

QY 162 GLGSGQTSRPGYGPQQTSRGGGLGQGGAGAAAAAAGGAGCGGYYGCG-LGSGQTS 220
DB 452 GTGGGAGGAGGAGGAGADNPTGIGTGGDGTGGAAGAGGAGGAGGAGGAGGAGG 511

QY 221 RGLGSGQGGAGAAAAA-----AAGGAGCGYGG-----GSQTSRGGGLG 263
DB 512 VGGAGCGGSDGAGGAGADADPGATGTTGAGGAGGAGGAGGAGGAGGAGGAGGAG 571

QY 264 GGGAGAAAAA-----AAGGAGCGYGGGLGSGT--SGRGLGQGGAGAAAAA 313
DB 572 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 631

QY 314 AAAGGAGCGYGG-LGSGQTSRGGYGPQQTSRGGGLGQGA-----CA 357
DB 632 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 688

QY 358 AAAAAAAGG-----AGCGYGGGLGSGT--SGRGLGQGGAGAAAA 398
DB 689 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 748

QY 399 AAAAAAGGAGCGYGG-LGSGQTSRGGGLGQ-----GAGAAAAA-----AAA 442
DB 749 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808

QY 443 GGA-GGGYGG-----GSQTSRGGGLGQGGAGAAAAAAGGAGCGGYYGGLGSGT 497
DB 809 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 866

QY 498 GPGG-YGPGQTSRGGGLGQGGAGAAAAAAGGAG-----GCGYGGGLGSGQTSR 552
DB 867 GGGNAGTCAGDPKCTGCTG-GTGGSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 923

QY 553 GGGAGAAAAAAGGAGCGYGGGLGSGQTSRGGGLGQGGAGAAAAAAGGAG----- 609
DB 924 NTDGLSSATSGTGGTGGGKGTGGAGDSDSAGGTGGTGGAGGAGGAGGAGGAGGAG 983

QY 610 ---GCGYGGGLGSGQTS-----RGLGQGG-----AGAAAAAAGGAGCGY 652
DB 984 GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1043

QY 653 GGLG-----SQTSRGGYGG 667
DB 1044 GGAGISFNSNGSGTGGTGGVG 1065

RESULT 12
Q8VIZO PRELIMINARY; PRT; 1715 AA.
AC Q8VIZO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
```

```
DE PE_PGRS family protein.
GN MT3612.1.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kelson J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE007163; AAK47971.1; -.
DR TIGR; MT3612; -.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR ProSITE; PS00583; PFKB_KINASES_1; UNKNOWN_4.
SQ SEQUENCE 1715 AA; 133832 MW; B857B2A774C771B0 CRC64;

Query Match      35.8%; Score 1268.5; DB 16; Length 1715;
Best Local Similarity 40.5%; Pred. No. 2.5e-61;
Matches 343; Conservative 33; Mismatches 282; Indels 188; Gaps 36;

QY 6 GGOQMGRGSMASGRGLGSGQ-----GAGAAAAA-----AA 35
DB 227 GGGAGGAGGAGGTGAGAGRAELLFGAGGAGGAGTGGPCATGCTGGHGVGGDGLWLP 286

QY 36 AAAGGA-GGGYGGGLGSGQ--TSRGGGLGQ-GAGAAAAAAGGAGCGGYYGGLGSG 91
DB 287 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 346

QY 92 TSGR-----GGLGQGGAGAAAAAAGGAGQ-----GCGYGGGLGSGQTSRGG 136
DB 347 GTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 406

QY 137 GGGG-AGAAAAAAGGAG-GGGYGGGLGQ-----GTSRGGYGPQQTSRGGGLG 187
DB 407 GCGAAGAGADAPASTGLTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 466

QY 188 GCGAGAAA-----AAAAAAGGAGCGYGG-----GSQTSRGGGLG 226
DB 467 GSGADNASGIGADGGAGGTGGNAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 526

QY 227 QCAGAAAAAAGG-----AGCGYGGGLGQ-----GTSRGGGLGQGGAGAAAAA 278
DB 527 AGSAGTATATCATGTTGTSRGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585

QY 279 GAGGAGGAGGGLGSGQTSRGGGLGQGGAGAAAAAAGGAG-GGGYGGGLGSG----- 330
DB 586 GADNPTGIGTGC-GTGGKAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643

QY 331 GTSGRGGYGPQQTSRGGGLGQGA-----GAAAAAAGGAG-GGGYGG 376
DB 644 GTNGSGGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 703

QY 377 -----LGSGQTSRGGGLGQ-GAGAAAAAAGGAG-----GCGYGGGLGSGQTSR 424
DB 704 AVGSVGNAGIGTGTGGVGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 761

QY 425 GLGQGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 484
DB 762 GTNGSG-----GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 815

QY 485 GCGYGG-----GSQTSRGGYGG--PGQTS-----RGLGQGGAGAAAAA 530
DB 816 TGGTGGVVGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 875
```



```
QY 235 AAAAAAGG-AGQGGYGGG------GTSRGLGGG--AGAAAAAAA 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 GGTGGAGTGGVGTGGSDGVNYSAGADGHPGGTGGVGTGGKGGDGDGGAAPNV 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 278 AGAGQGGYGGGSGGTSRGLGGGAGAAAAAAG-GAQGGYGGGSG-GTSGP 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 AGSQGGAGG--GTGGVGGNGRIDGADGATAGAQDGGAGGKGGGTGP 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 GGYGP-----GQTSRGLGGG-----AGAAAAAAA 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 815 GGAGPACTTGSQAGGNGSGCTGGDPDGGNGANGSVFTNNGIGGNGGNGAGPSGAG 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 365 AAGGA-----GQGGYGGGSGGTSRGLGGGAGAAAAAAAAGGA 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 875 GSGAGCTFGATGSSSIHVNNGNGNGDHALSCNGAAGNGGNGSLRGSGAG 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 408 QGGYGGGSGGTSRGLGGGAGAAAAAAGAGAGGGYGG-----LGSQGTSG 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 GHGNGGNASRGMGGDGTGGAGNAGQIGNGGAGNGDGTGSDGNPGAITGSGRGG 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 461 RGLGGGAGAAAAAAGAGAGGGYGGGSGGT-----SGPGYGPQQT 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 DGGVGGGSGVAGDAGDGRGA--GGTGTGLRGTGATGCTFDAGADGHGNGGTG 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 GRGLGGGAGAAAAAAA-----GGAGGGYGGGSGGTSRGLGGGA-G 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 GVGGTGGAGGGGNGCAGGKALSPTGNGSQAGGAGGAGG--GTGTTGGDGRGAGH 1110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 AAAAAAAGGAGGQGGYGGGSGGTSRGLGGGAGAAAAAAGAGGGYGGG 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1111 TLFSSLAGTGTGNGGTGTG--GTGAGGAGGTGTLGATGATGAAGRAGNGVGGSG 1168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 SQGTS-CRGLGGGAGAAAAAAGAGAGGGYGGG-----SQGTSRGGY-GPCQQT 672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1169 GLGSFPGGTGGMG-GAGGTSTVSAGDGGGFGGDLDASSGNGGDDGGDFRT 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 673 SG 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 AG 1229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 ; Search time 35.254 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700
Sequence: 1 MASMTGGQMGRIIRIRGG.....GLSGCDVLIQALLGHHHHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2700	100.0	528	22	AA82611 Spider recombinant
2	2319	85.9	646	18	AAW27178 Nephila clavipes s
3	2196	81.3	718	19	AAW53346 Nephila clavipes s
4	2196	81.3	718	21	AAV59070 N. clavipes spider
5	2182	80.8	718	12	AAK14308 N.clavipes draglin
6	1999	74.0	651	20	AAV40097 Spider silk protei
7	1999	74.0	651	23	AAU11781 Spider natural sil
8	1996	73.9	1255	23	AAU11781 N. clavipes spider
9	1996	73.9	1880	23	AAW50037 N. clavipes spider
10	1970	73.0	989	23	AAW50038 N. clavipes spider

11	1944	72.0	818	23	AAU11797 Dragline protein 1
12	1944	72.0	1626	23	AAU11796 Dragline protein 1
13	1919	71.1	630	23	AAW50042 N. clavipes spider
14	1919	71.1	676	23	AAW50047 N. clavipes spider
15	1914	70.9	809	23	AAU11793 Dragline protein 1
16	1914	70.9	1617	23	AAU11794 Dragline protein 1
17	1883.5	69.8	606	16	AAW59055 Spider dragline va
18	1883.5	69.8	606	20	AAV40102 Polymer of an anal
19	1879.5	69.6	606	20	AAV40101 Polymer of an anal
20	1871.5	69.3	604	16	AAW59057 Spider dragline va
21	1858.5	68.8	606	16	AAW59053 Spider dragline va
22	1858.5	68.8	606	20	AAV40100 Polymer of an anal
23	1784.5	66.1	691	22	AAW59055 Spider recombinant
24	1711.5	63.4	681	22	AAW59055 Spider recombinant
25	1623.5	60.1	777	23	AAW50046 N. clavipes spider
26	1588.5	58.8	773	23	AAW50045 N. clavipes spider
27	1587.5	58.8	364	23	AAW50043 N. clavipes spider
28	1463	54.2	655	23	ABW76672 Protein related to
29	1344	49.8	831	16	AAW50035 PMISS1 Misp spider
30	1234	45.7	615	20	AAV40039 Spider silk protei
31	1209.5	44.8	271	23	AAW50044 N. clavipes spider
32	1144	42.4	595	12	AAW14309 N. clavipes draglin
33	1144	42.4	595	19	AAW53347 Nephila clavipes s
34	1144	42.4	595	21	AAV59071 N. clavipes spider
35	1042.5	38.6	230	23	AAW50035 N. clavipes spider
36	1008	37.3	291	22	AAW50035 Spider recombinant
37	973.5	36.1	465	22	AAW70188 Peptide dendrimer
38	972	36.0	1177	9	AAW50040 SIIPIII protein com
39	972	36.0	1177	11	AAW50307 SLP III (Silk-fibr
40	972	36.0	1177	17	AAW51105 Silk like protein
41	972	36.0	1177	18	AAW26342 Silk-like protein
42	972	36.0	1177	21	AAW78277 SIIPIII amino acid
43	972	36.0	1177	22	AAW72725 Repetitive protein
44	972	36.0	1177	22	AAW63995 SIIPIII protein seq
45	972	36.0	1178	14	AAW41007 Silk-like protein

ALIGNMENTS

RESULT 1
AAW82611
ID AAW82611 standard; Protein; 528 AA.
XX AAW82611;
AC AAW82611;
XX 02-OCT-2001 (first entry)
DT
DE Spider recombinant silk protein pETNCDs.
DE
XX Spider; orb-weaver; silk protein; pETNCDs; structural protein;
KW purification; fibre; spinning.
KW
XX Nephila clavipes.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 417 /note= "encoded by TAT"
FT Misc-difference 427 /note= "encoded by CGA"
FT Misc-difference 522 /note= "encoded by GAG"
FT
XX WO200153333-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 01-NOV-2000; 2000WO-US30086.
PF
XX
XX 20-JAN-2000; 2000US-0490291.
PR
XX
XX (WELL/) MELLO C M.
PA (ARCI/) ARCIDIAONO S.

XX AC AAU11781;
XX XX
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Spider natural silk protein Spidroin 1.
XX XX
XX KW Orb-weaving spider; spidroin 1; silk-like protein; SLP; transgenic plant;
XX KW promoter; 5' terminator; fabric production; material construction; rope;
XX KW surgical suture; flexible tie down; electrical component; implantation.
XX OS
XX XX Nephila clavipes.
XX XX WO200190389-A2.
XX XX
XX PD 29-NOV-2001.
XX XX
XX XX 24-MAY-2001; 2001WO-US16937.
XX XX
XX PR 25-MAY-2000; 2000US-206968P.
XX XX
XX FA (DUPO) DU PONT DE NEMOURS & CO E I.
XX XX
XX PT Yang JG;
XX XX
XX DR WPI; 2002-106209/14.
XX XX
XX PT Producing silk-like proteins in a green plant, useful in fabrics and
XX PT material construction, comprises providing a silk-like protein
XX PT expression cassette to a green plant -
XX XX
XX PS Disclosure; Page 69-71; 93pp; English.
XX XX
XX CC The invention relates to a method of producing silk-like proteins (SLP)
XX CC in a green plant by providing a green plant containing an SLP expression
XX CC cassette. The green plant contains the expression cassette P-SLP-T where
XX CC P is a promoter for driving the expression of an SLP transgene and T is a
XX CC 5' terminator. The method is useful for producing silks and silk-like
XX CC proteins in green plants. The silks and silk-like proteins may be used in
XX CC fabrics or in material construction, such as rope, surgical sutures,
XX CC flexible tie downs for certain electrical components, or as a
XX CC biomaterial for implantation. The method allows for more cost effective
XX CC production of silk not obtained from natural or microbial sources.
XX CC The present sequence is orb-weaving spider spidroin1, a natural silk
XX CC protein whose sequence is used as the basis for the SLP proteins used in
XX CC the method of the invention.
XX SQ Sequence 651 AA;

Query Match 74.0%; Score 1999; DB 23; Length 651;
Best Local Similarity 68.2%; Pred. No. 3.8e-130;
Matches 435; Conservative 5; Mismatches 10; Indels 188; Gaps 15;

QY 6 GQQQMGTRIRYGGGLGCGGAGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAA 65
DB 22 GQGGAGQ---GGYGGGCGGAGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAA 78
QY 66 AAAGGAGGGYGGGLGSGQAGRCGCGAG---AAAAAGGGYGGGLG---CGAGCGA-GA 119
DB 79 AAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGA 138
QY 120 AAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGA 166
DB 139 AAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGA 198
QY 167 GLGSGAGRCGCGAGGAGCA-----185
DB 199 GLGSGAGRCGCGAGGAGCAAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGGAGGAG 258
QY 186 -----AAAAAGGAGGGYGGGLGSGQAGRCGCGAGGAGGAGGAGGAGGAGGAGGAG 232
DB 259 GSGAGRCGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGGAGGAGGAGGAGGAGGAG- 317

QY 233 AAAAGAGGGGGLG-----GGCGAGAAAAAGGAGGGYGGGLGSGQAGRC---GGCGAGA-AAA 284
DB 318 AAAAGGAGGGGLGCGGAGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGGAGGAGAAA 377
QY 285 AAGGAGGGYGG---QGAGQG-----GYGGGLGSGQAGRCGCGAGGAGA 324
DB 378 AAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGRCGCGAGGAGGAGGAGGAGGAGGAGA 437
QY 325 AA-----AAAAAGGGAGQ-----AAAAAGGAGQ-----336
DB 438 AAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAA 497
QY 337 -----GGLGCGCGAGAAAAAAGGAGGGYGGGLGCGGAG 365
DB 498 AAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGCGGAG 557
QY 366 QGAGAAAAAAGGAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGCGGAG 425
DB 558 QGAG-----AAAAAGGGYGGGLGSGQAGRCGCGAGAAAAAAGGAGGGYGGGLGCGGAG 613
QY 426 GAGGLGCGGAGAAAAAGGAGGGYGGVGGSGASAAASAAAS 463
DB 614 GGGGLGCGGAGAAAAAGGAGGGYGGVGGSGASAAASAAAS 651

RESULT 8
AAM50037
ID AAM50037 standard; Protein; 1255 AA.
AC AAM50037;
XX 18-SEP-2002 (first entry)
XX XX
XX DE N. clavipes spidroin synthetic homologue SOISOL protein.
XX KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; SOISOL.
XX OS Synthetic.
XX PN DE10113781-A1.
XX PD 13-DEC-2001.
XX XX
XX PF 21-MAR-2001; 2001DE-1013781.
XX PR 09-JUN-2000; 2000DE-1028212.
XX PR 24-OCT-2000; 2000DE-1053478.
XX XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX Scheller J, Conrad U, Grosse F, Guehrs K;
XX WPI; 2002-123561/17.
XX N-PSDB; ABL61038.
XX PT New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX PT wounds, comprises modules that encode repeating units of spidroin
XX PT proteins -
XX PS Claim 22; Page 29-34; 88pp; German.
XX XX
XX CC This invention describes a novel DNA sequence, encoding a synthetic
XX CC spider silk protein, comprising modules, each comprising a group of
XX CC sequentially arranged oligonucleotides, each oligonucleotide encoding
XX CC a repeating unit of a spidroin protein. The synthetic protein has at
XX CC least 84% homology with the Nephila clavipes spidroin protein and is used
XX CC to produce synthetic fibres, films and/or membranes, particularly: (i)
XX CC for medical use, especially to close wounds and/or to support or cover
XX CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX CC (iii) as filters. The synthetic proteins are very similar to native
XX CC spider silk proteins; can be prepared on a large scale and can be spun to
XX CC fibres with excellent mechanical properties (strength and elasticity).
XX CC Also they retain water solubility after long-term boiling in aqueous

Best Local Similarity 83.5%; Pred. No. 1.3e-124;					
Matches 404; Conservative 4; Mismatches 16; Indels 60; Gaps					
Qy	6	GGQQMGRIRINGVGLGGQGAGGACAGAAAAAAGGACGGYGGILGSQAGRGGGGAGAA 65			
		:			
Dd	37	GGCQNGQ---GGYGGLGGGAGGAG--AAAAGAAGCGGYGGILGSQAGRGGGGAGAA 92			
Qy	66	AAAGAGAGGGYGGILGSQAGRGGILGGGAGAAAAAAGGVGGGLGGGAGGAGAAAAAAG 125			
		:			
Dd	93	AAAGAGAGGGYGGILGSQAGRGGILG-----GGACAAAAAAG 129			
Qy	126	GAGGGYGGILGSQAGRGGSGGGGAGAAAAAAGAGGGYGGILGSQAGRGGILGGGAGAA 185			
		:			
Dd	130	GAGGGYGGILGSQAGR---GGCAGAAAAAAGAGGGYGGILGSQAGRGGILGGGAG-- 185			
Qy	186	AAAAAAGAGAGGGYGGILGGGAGGGYGGILGSQAGRGGILGGGAG--AAAAAGGACGGGL 244			
Dd	186	AAAAAAGAGAGGGYGGILGGGAGGGYGGILGSQAGRGGILGGGAGAAAAAAGGAGGGGL 245			
Qy	245	G-----GGCAAAAAAAGAGGGYGGILGSQAGRGGGGAGAAAAAAGGAGGGY---GGQ 297			
		:			
Dd	246	GGCAGGAGGAAAAAAGAGGGYGGILGSQAGRGGGGAGAAAAAAGGAGGGYGGILGGCQ 305			
Qy	298	GAGGGYGGILGSQAGRGGILGGGAGAAAAAAGAGAGCGGLG----GGAGAAAAAAGG 353			
		:			
Dd	306	GAGGGYGGILGSQAGRGGILGGGAG--AAAAAAGAGGGYGGILGGGAGGAGAAAAAAGG 363			
Qy	354	AGCGGLGG---QGAGG--ACAAAAAAGGVROGGYGGILGSQAGRCGCAGAGAAAAA 408			
		:			
Dd	364	AGCGYGGILGSQAGRGGILGGGAGAAAAAAGAGGGYGGILGSQAGRGGGGAGAAAAA 423			
Qy	409	AGGAGCGGTGTLGGGQGVGAGLGGGAGAAAAAV--GAGGGYGGVGVS-----GASAA 458			
		:			
Dd	424	AGGAGGGYGGILGSQAGRGGILGGGAGAAAAAAGAGGGYGGILGSQAGRCGCAGAA 483			
Qy	459	SAAA 462			
		:			
Dd	484	AAAA 487			
 RESULT 15					
AAU11793					
ID	AAU11793 standard; protein; 809 AA.				
XX					
AC	AAU11793;				
XX					
DT	26-MAR-2002 (first entry)				
XX					
DE	Dragline protein 1 analogue DP-1B 8mer.				
XX					
Kw	Orb-weaving spider; silk-like protein; SLP; transgenic plant;				
Kw	promoter; 5' terminator; fabric production; material construction; ro-				
Kw	surgical suture; flexible tie down; electrical component; implanta-				
Kw	Dragline protein 1; DP-1B 8mer.				
XX					
OS	Nephila clavipes.				
OS	Synthetic.				
XX					
PN	WO200190389-A2.				
XX					
XX	29-NOV-2001.				
PF	24-MAY-2001; 2001WO-US16937.				
XX					
PR	25-MAY-2000; 2000US-206968P.				
XX					
PA	(DUPO) DU PONT DE NEMOURS & CO E I.				
XX					
PI	Yang JG;				
XX					
DR	WPI; 2002-106209/14.				
XX					
PT	Producing silk-like proteins in a green plant, useful in fabrics and				

PT material construction, comprises providing a silk-like protein
XX expression cassette to a green plant -
PS Disclosure; Page 74-76; 93pp; English.
XX

XX The invention relates to a method of producing silk-like proteins (SLP)
CC in a green plant by providing a green plant containing an SLP expression
CC cassette. The green plant contains the expression cassette P-SLP-T where
CC P is a promoter for driving the expression of an SLP transgene and T is a
CC 5' terminator. The method is useful for producing silks and silk-like
CC proteins in green plants. The silks and silk-like proteins may be used in
CC fabrics or in material construction, such as rope, surgical sutures,
CC flexible tie downs for certain electrical components, or as a
CC biomaterial for implantation. The method allows for more cost effective
CC production of silk not obtained from natural or microbial sources.
CC The present sequence is the silk protein Dragline protein (DP) 1
CC synthetic variant DP1B 8mer (i.e. 8 copies of DP1B) which is used as
CC an SLP construct in the method of the invention.
XX

SQ Sequence 809 AA;

Query Match 70.9%; Score 1914; DB 23; Length 809;
Best Local Similarity 65.7%; Pred No. 3.3e-124;
Matches 425; Conservative 1; Mismatches 27; Indels 194; Gaps 17;

QY	6	GCQQMGRIRI-----RGYGLGGQAGGAGAGAAAAAGAGGQGGYGGGLGSGQAGRGGGQ	60
Db	125	GGQAGAGAAAAAGGAGGQGLGSGQAGGAG-AAAAAGAGGQGGYGGGLGSGQAGRGGGQ	183
QY	61	AGAAAAAG-----GAGQGGYGGGLGSGQAGRGGLGGQAG--AAAAAGGVGQ	105
Db	184	AGAAAAAGGAGGQGGYGGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAGAAAAAGGAGQ	243
QY	106	GGGLGQAGAGAGAAAAAGGAGGQGGYGGGLGSGQAGRGSGGQGGAGAAAAAG-----	158
Db	244	GGGLGSGQAGGAGAAAAAGGAGGQGGYGGGLGSGQAGR---GGQAGAAAAAGGAGGQGGY	300
QY	159	-----GAGQGGYGGGLGSGQAGRGGLGGQAG-----	185
Db	301	GGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAGAAAAAGGAGGQGGGLGSGQAGGAGAAA	360
QY	186	-----AAAAAGGAGGQGGYGGGLGSGQAGGQGGYGGGLGSGQ	218
Db	361	AAAGGAGGQGGYGGGLGSGQAGRGQGGAGAAAAAGGAGGQGGYGGGLGSGQ	420
QY	219	GAGRGGLGSGQAG--AAAAAGGAGGQGG--GQAGAAAAAGGAGGQGGYGGGLGSGQ	272
Db	421	GAGRGGLGSGQAGAAAAAGGAGGQGGGLGSGQAGGAGAAAAAGGAGGQGGYGGGLGSGQ	480
QY	273	GRGGGAGAAAAAGGAGGQGGYGG--GQAGQGGYGGGLGSGQAGRGGLGGQAGAA-----	325
Db	481	GRGGGAGAAAAAGGAGGQGGYGGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAGAAAAA	540
QY	326	-----AAAAAGGAGGQGG-----GQAGAGAAAAAGGAGQ	356
Db	541	AGGAGQGGGLGSGQAGGAGAAAAAGGAGGQGGYGGGLGSGQAGRGQGGAGAAAAAGGAGQ	600
QY	357	-----GGLGGQAG-----QGAGAAAA	373
Db	601	GGYGGGLGSGQAGGQGGYGGGLGSGQAGRGGLGGQAGAAAAAGGAGGQGGGLGSGQAGGAG	660
QY	374	AAAAAGGVROGGYGGGLGSGQAGRGQGGAGAAAAAG-----GAGQGGTGGL	420
Db	661	AAAAAGGAGGQGGYGGGLGSGQAGRGQGGAGAAAAAGGAGGQGGYGGGLGSGQAGGQGGY	720
QY	421	CGQGVGAGGLGQAGAAAAV---GAGQGGY--GVGSGASASAAA	462
Db	721	GSQAGRGGLGGQAGAAAAAGGAGGQGGGLGSGQAGGQAGAAAAA	767

GenCore version 5.1.3
Copyright (C) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 : Search time 12.1566 Seconds
(without alignments)
1277.936 Million cell updates/sec

Title: US-09-490-291-8

Perfect score: 2700

Sequence: 1 MASMTGGQMGIRIRYGG.....GLSGCDVLIQLLHHHHHH 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	84.2	747	3	US-09-034-177-3
2	2196	81.3	718	1	US-08-425-069-2
3	2196	81.3	718	2	US-08-317-844B-2
4	1999	74.0	651	4	US-08-556-978B-19
5	1999	74.0	651	4	US-09-247-806-1
6	1883.5	69.8	606	4	US-08-556-978B-23
7	1883.5	69.8	606	4	US-09-247-806-8
8	1873.5	69.3	604	4	US-09-247-806-6
9	1871.5	69.3	604	4	US-08-556-978B-63
10	1858.5	68.8	606	4	US-08-556-978B-21
11	1858.5	68.8	606	4	US-09-247-806-4
12	1343.5	49.8	832	1	US-08-209-747-2
13	1343.5	49.8	832	1	US-08-458-298-2
14	1144	42.4	595	1	US-08-425-069-4
15	1144	42.4	595	2	US-08-317-844B-4
16	972	36.0	1136	4	US-08-806-029-9
17	972	36.0	1177	1	US-07-609-716-31
18	972	36.0	1177	1	US-08-175-155-29
19	972	36.0	1177	1	US-08-477-509B-64
20	972	36.0	1177	2	US-08-707-237A-35
21	972	36.0	1177	3	US-08-482-085B-64
22	972	36.0	1177	4	US-08-475-411A-31
23	972	36.0	1177	4	US-08-478-029A-36
24	972	36.0	1177	4	US-09-444-791A-83
25	970	35.9	714	4	US-08-556-978B-61
26	970	35.9	714	4	US-09-247-806-10
27	962.5	35.6	1059	1	US-08-175-155-48

28	962.5	35.6	1059	2	US-08-707-237A-54	Sequence 54, Appl
29	962.5	35.6	1059	4	US-08-806-029-10	Sequence 10, Appl
30	962.5	35.6	1101	1	US-08-477-509B-83	Sequence 83, Appl
31	962.5	35.6	1101	3	US-08-482-085B-83	Sequence 83, Appl
32	962.5	35.6	1101	4	US-09-444-791A-83	Sequence 83, Appl
33	929	34.4	1332	1	US-07-609-716-41	Sequence 41, Appl
34	929	34.4	1332	4	US-08-475-411A-41	Sequence 41, Appl
35	929	34.4	1332	4	US-08-478-029A-41	Sequence 41, Appl
36	922.5	34.2	1038	1	US-07-609-716-36	Sequence 36, Appl
37	922.5	34.2	1038	4	US-08-475-411A-36	Sequence 36, Appl
38	922.5	34.2	1038	4	US-08-478-029A-36	Sequence 36, Appl
39	921.5	34.1	766	1	US-08-175-155-53	Sequence 53, Appl
40	921.5	34.1	766	1	US-08-477-509B-88	Sequence 88, Appl
41	921.5	34.1	766	2	US-08-707-237A-61	Sequence 61, Appl
42	921.5	34.1	766	3	US-08-482-085B-88	Sequence 88, Appl
43	921.5	34.1	766	4	US-09-444-791A-88	Sequence 88, Appl
44	921.5	34.1	979	1	US-08-477-509B-89	Sequence 89, Appl
45	921.5	34.1	979	3	US-08-482-085B-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1

US-09-034-177-3
; Sequence 3, Application US/09034177
; Patent No. 6127146
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,177
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0486 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: GI 1174414
US-09-034-177-3

Query Match 84.2%; Score 2273; DB 3; Length 747;
Best Local Similarity 71.0%; Pred. No. 2.7e-160;

Db 343 GPGYGPSSASAAAAAGPGQCGPGYGPQCGPSGPGSASAAAAAAGPGYGPQCG 402
QY 359 LGCGGAGQ-----CAGAAAAAAGAGVROGGYGGGLGSGAG-----RGCGAG-AAAA 407
Db 403 PGYAPGQCGPGSPGPGSASAAAAAAG--PGYGP-GQCGPGYAPGQCGPSGPGSAAAA 459
QY 408 AAGGAGGGTG-----GLGGGV-----CAGGLG-----GOGAGAAAAGAGGG 447
Db 460 AAASAGPGYGPAAQPGSPGPGIAASASAGPGYGPAAQPGYGPAAQPGYGPAAQPGYGP 519
QY 448 YGVGSGASASAAAAAASRLSPQASRVSAVSNLVASGPTNSAALSSTISNVVYSGAGSN 507
Db 520 YG----PGSQASAAASRLASPDGSAVSNLVSSGPTSSAALSSVSNVYSGAGSN 575
QY 508 PGLSGCDVLIQAL 520
Db 576 PGLSGCDVLIQAL 588

RESULT 15

US-08-317-844B-4
; Sequence 4, Application US/0831784B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hanman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-844B-4

Query Match 42.4%; Score 1144; DB 2; Length 595;
Best Local Similarity 51.1%; Pred. No. 3.4e-77;
Matches 313; Conservative 25; Mismatches 139; Indels 136; Gaps 30;
QY 17 GYG-----GLGGGAGQ-----GAG-AAAAAAGAGAGGGY-----GLGSGAGR-- 56
Db 3 GYGPGQCGPGYGPQCGPGSPGPGSASAAAAAAGPGYGPQCGPGYGPQCGPGRYG 62
QY 57 -----GQCGAGAAAAAGAGAGGGYGGGLGSGAGRGGLG-GQ-----CAGAAAAAGGV---- 103

Db 63 PGQCGSPGPGSAAAAAAGSGQCGPGYGPGRQCGPGYGGQCGPGPGSAAAAAASASAE 122
QY 104 -GQGGGLGCGAGCGAGAGAGAGAGGGY-----GLGSGAGRGSGGGQ--AGAAAAA 157
Db 123 SGQCGPGYGPQ-----QCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGY 171
QY 158 GAGQ---GGY-----GLGSGAGRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 210
Db 172 SGPGQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPG 231
QY 211 GYGGLGSGAGRGGLGCGAG 261
Db 232 GY-----GPGQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPG 285
QY 262 GYG-----GLGSGAG--RGCGAG--AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 315
Db 286 GYGPGQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPG 342
QY 316 GLGGGAGAGAAAAAAG 358
Db 343 GPGYGPAGSASAAAAAAGPGQCGPGYGPQCGPGYGPQCGPGYGPQCGPGYGPQCGPG 402
QY 359 LGCGGAGQ-----CAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
Db 403 PGYAPGQCGPGSPGPGSASAAAAAAG--PGYGP-GQCGPGYAPGQCGPGSPGPGSAAAA 459
QY 408 AAGGAGCGGTG-----GLGGGV-----CAGGLG-----GOGAGAAAAGAGAG 447
Db 460 AAASAGPGYGPAAQPGSPGPGIAASASAGPGYGPAAQPGYGPAAQPGYGPAAQPGYGP 519
QY 448 YGVGSGASASAAAAAASRLSPQASRVSAVSNLVASGPTNSAALSSTISNVVYSGAGSN 507
Db 520 YG----PGSQASAAASRLASPDGSAVSNLVSSGPTSSAALSSVSNVYSGAGSN 575
QY 508 PGLSGCDVLIQAL 520
Db 576 PGLSGCDVLIQAL 588

Search completed: December 18, 2002, 16:23:13
Job time : 16.1566 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1999	74.0	651	10	US-09-861-597-1	Sequence 1, Appli	
2	1883.5	69.8	606	10	US-09-861-597-8	Sequence 8, Appli	
3	1873.5	68.4	606	10	US-09-861-597-6	Sequence 6, Appli	
4	1858.5	68.8	606	10	US-09-861-597-4	Sequence 4, Appli	
5	970	35.9	714	10	US-09-861-597-10	Sequence 10, Appli	
6	915.5	33.9	529	10	US-09-861-597-2	Sequence 2, Appli	
7	679	25.1	357	10	US-09-864-761-35807	Sequence 35807, A	
8	649	24.0	1894	12	US-10-052-586-97	Sequence 97, Appli	
9	557	20.6	4679	10	US-09-804-898-2	Sequence 2, Appli	
10	534	19.8	3907	9	US-10-029-217A-24	Sequence 24, Appli	
11	529	19.6	450	10	US-09-812-382-6	Sequence 6, Appli	
12	524	19.4	720	10	US-09-756-071B-20	Sequence 20, Appli	
13	523.5	19.4	283	10	US-09-864-761-36720	Sequence 36720, A	
14	511	18.9	2211	9	US-10-096-961-1	Sequence 1, Appli	
15	500.5	18.5	201	10	US-09-848-990-22	Sequence 22, Appli	
16	500.5	18.5	201	10	US-09-760-364-14	Sequence 14, Appli	
17	499.5	18.5	200	10	US-09-798-584-18	Sequence 18, Appli	
18	499.5	18.5	200	10	US-09-967-624-19	Sequence 19, Appli	
19	499.5	18.5	200	10	US-09-938-667-18	Sequence 18, Appli	

;
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212

;
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908
;
Query Match 24.0%; Score 649; DB 12; Length 1894;
Best Local Similarity 32.8%; Pred. No. 5.4e-32;
Matches 216; Conservative 12; Mismatches 230; Indels 200; Gaps 19;
Qy 26 ACQGAGAAAAAGGAGGGGGLGSGAGRGQGQAGAAAAAGGAGGGGGLGSGAG 85
Db 719 AGGGAAATTTAAAGGC-----TGAAAAAGAGAACCCAGAGGAGTAAGAA 768
Qy 86 R-----GGLGGQAGAAAA-----GGVGGGLGSGAGQGA 117
Db 769 ATTGAAACCAAGGCACAAAAATTTTACTTTCTTTGGAGAGAG--CTGA 825
Qy 118 GAAAAAGGAGGGGGLGSGAGRGGGGGGAGAG-----AAAAAGGAGGGYGG 167
Db 826 GGAAGAGAGGAGGAAGTAATCGAGTTAGTCAGACATGAAGGGCAAAAGCAAGTAG 885
Qy 168 LGSQG-----AGRGGLGGQAGAAAA 190
Db 886 TCATGACTTGTAAAGGATGCCACATCTCAGTTCTGTTCAGTTGTAGAAAGTGA 945
Qy 191 AGGAG-----QGGYGGGLGSGAG--GQGGYGGGLGSGAG----- 221
Db 945 AGGTGATGCCACGAGATTTAGTTGATGATGAGAGATGAAGTGCAGACATGATGA 1005
Qy 222 ----RGLGGQAGAAAAAGGAGGGGLGSGAG-----AAAAAGGA 259
Db 1006 TATTGATGTTGATGAAACCAACCTGATGAGAGAAAGAAATTGCAAAAAATTTAAAAAGCA 1065
Qy 260 -----GQGGYGGGLGSGAGRGQGQAGAAAAAGGAGGG 293

Qy 241 -QGG-----LGGGAGAAAAAG--GAGQGGYGLG 268
Db 1350 CGGGCTCCACGCCCCCGTGTCTCCACCCCTCGAGCGCTCACTGCTCAGCAGCGGG 1409
Qy 269 SGGAGR-----GGGAGAAAAAGAGGQGGYGGGA----- 299
Db 1410 ATGAAAACCTCCACCCCGGGACACCTTTGGTGTAGATGGTACATCATCTCTACGCGAGC 1469
Qy 300 -----GQGGYGLGSGQ-----AGRGGLGG-- 319
Db 1470 TCACCTCGAGCCCTCGCCACTGCAGATCCTCGTGAAGGAGGAGGCCCGCGCGCGGT 1529
Qy 320 -----QGAGAAAAAAGAGAGQ-----GLGQGGAGAAAAAGGA-CGGGLGG 361
Db 1530 CCGTGTCTGAGCCCTGGGGGGCGGAGCTAGAGGGCGGCACAAGGACCATGTC 1589
Qy 362 QGAGGAGAAAAAAGAGGVRGGYGLGSGA-----GRGGGAGAAAAAGGAGGG 416
Db 1590 TGCAGGAAAGACAAAGCAGATCGA--GGCGTGTAGCGCATGCTCCGCGAGAAGCAGCAG 1648
Qy 417 -TGLGGGGVG-----AGGLGGQ--AGAAAAVAGGQ 446
Db 1649 CTGTTGAGCGCTCAACCTGAGCTGAGCAGGAGAGAGCGAG 1691
RESULT 11
US-09-812-382-6
; Sequence 6, Application US/09812382
; Patent No. US20010034050A1
; GENERAL INFORMATION:
; APPLICANT: Chilkoti, Ashutosh
; TITLE OF INVENTION: Fusion Peptides Isolatable by Phase Transition
; FILE REFERENCE: 4176-101
; CURRENT APPLICATION NUMBER: 2001-03-20
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/190,659
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 450
; TYPE: PRT
; ORGANISM: ELP[V5A2G3-90]
US-09-812-382-6
Query Match 19.6%; Score 529; DB 10; Length 450;
Best Local Similarity 35.9%; Pred. No. 2.2e-25;
Matches 165; Conservative 16; Mismatches 234; Indels 44; Gaps 16;
Qy 17 YGGLGGQGGAG-QGAGAAAAAAGAGGQGGYGLGSGQAGRGQGGAGAAAAAGGAGGG 75
Db 15 GVPGVGVPVPGAGVPGAGVPGGGVPG--GGV----PGGGVPGVPGVPGVPGVPG 68
Qy 76 YG--GLGSGQAGRGLGQGGAGAAAAAGCGVGGGLGGQGGAGGAGAAAAAGGAGGGY 133
Db 69 VGVPGVGPGAGVPGAGVPGGG--VPGGGVPGGGVPGV--PGVPGVPGVPGVPGVPG 125
Qy 134 --GLGSGQAGRGSGGQGGAGAAAAAGAGGGY--GLGSGQAGRGLGSGGAGAAAAA 189
Db 126 VPGAGVPGAGVPGGGVPGGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGAG 185
Qy 190 AAGGAGGG----YGLGQGGAGQGGYGLGSGQAGRGLGQGGAGGAGGG 243
Db 186 VPGGVPGGGVPGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGGG 245
Qy 244 LGG---QGAGAAAAAAGAGGGY--GLGSGQAGRGGQGGAGAAAAAGGAGGGYGGQ 298
Db 246 VPGGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPG 305
Qy 299 AGQGGYGLGSGQAGRGLGQGGAGAAAAAAGGAGGGYGG---QGAGAAAAAAGGAG 355
Db 306 VPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPG 365

Qy 356 QGGLGGQGGAGGAGAAAAAAGGVRGGYGLGSGQAGRGQGGAGAAAAAAGGAGGG 415
Db 356 VPGVPGVPGV--PGAGVPGAGVPGGGVPGGGVPGGGVPGVPGVPGVPGVPGV 414
Qy 416 GTGGLGGQGGVAGGGLGGQGGAGAAAAAAGAGGQGGYGGVGG 454
Db 415 GVPGVGVPVPGVPGAGVPGAGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPG 450
RESULT 12
US-09-756-071B-20
; Sequence 20, Application US/09756071B
; Patent No. US20020052307A1
; GENERAL INFORMATION:
; APPLICANT: tryggvason, Karl
; Kallunki, Pekka
; Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
; STREET: 1100 Superior Ave, Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,071B
; FILING DATE: 08-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/663,147
; FILING DATE: 150-September 2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard, J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 20014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-861-5582
; TELEFAX: 216-241-1666
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-756-071B-20
Query Match 19.4%; Score 524; DB 10; Length 720;
Best Local Similarity 30.4%; Pred. No. 6.6e-25;
Matches 210; Conservative 4; Mismatches 204; Indels 272; Gaps 29;
Qy 19 GGLGGQGGAGGAGAAAAAAGAGGQGGYGLGSGQAGRGQGGAGAAAAAAGGAGGGYGG 78
Db 12 GGAGGTTGCCAGTCAATA-----GGTTACTTTATGAGTTGCTAACCTGCTGACGAG 65
Qy 79 -LGSQAGRG--LGGQAGAAAAAGVGGGLGGQGGAGGAGAAAAAAGGAGGGYGG 134
Db 66 AAGTTATGAGCACCAGGAGAGAAACCCCTTG----GTTGAGCCTGGAGAAAGGAGGTTG 121
Qy 135 -----LGSQAGRGSGGQGGAGAAAAAAGGAGGG----- 164
Db 122 ACCCTAACTGGAGGTGGAGGACCCCTGTTGTGACTCTCGACTGACTTGTCTCTCTT 181
Qy 165 -----YGLGSGQAGR--GGLGG-----QGAG-----A 185

Db 182 GATGTCCTTTAAGCCGAGCTGATTCGGCTGCTGCTGCTTATTTCTCAGTTAGCGCTCTTA 241
QY 186 AAAAAAGG-----AGGGYGLGGGAG----- 208
Db 242 AGATTGGCCCTCCCACTTTGAGGAGGGCGGGCTGCTCTACCTCTCTGTAATCTGCC 301
QY 209 -----QGGYGLGSGAGRGG---LGQGA-----GAAAAAG-- 237
Db 302 TGGACACCCCGGAGAGAGGAGGCTCCGGGAATCTCGACATTCAGGCAAGGCT 361
QY 238 ---GAGGGLGG---QGAGAAAA-----AGAGGG----- 262
Db 362 CCGGGCCGAGCCTCTGTGCCACACCTTGGCCCGGCGAGGTGTGCGCCTCTCGCT 421
QY 263 GYGLGSGAGRGGQAGAAAAAGGAGG-----GAG-----GAG- 300
Db 422 GCGAGGGGAGGCGGGGCTGCGGGAGCGATTTTCCACCCGGTTGTGCTGTGTGT 481
QY 301 -----QGGYGLGSGAGR-----GGLGGGAGAAAAAAGGAG--- 335
Db 482 TTGCTGCTCTGGAGG-GCTGGTCTCTTATTCACAGGTGAGTCACACCTGAAACA 540
QY 336 -QGG-----LGGGAGAAAAAAGGAGGGLGGGAGGGA-----GAAAAA 374
Db 541 CAGGCTCTCTCTGTCAGGACTGAGTCAGGTAGAGAGTGCATATAAACACCTGATCAA 600
QY 375 AAAAGGVRGGYGLGSGAGRG-----CQAGAA---AAAAGAGGGTGGGGGAG 428
Db 601 GAAAAAGGAAGGACACAGCGGAGCGAGAGTGAACCAACCGAGGCGCGGGCAGCG 660
QY 429 -----GLGGGQ--AGAAAAVAGGCG 447
Db 661 ACCCTGACAGCGAGACAGACTGAGCG 690

RESULT 13
US-09-864-761-36720
; Sequence 36720, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rankel, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36720
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-36720

Query Match 19.4%; Score 523.5; DB 10; Length 283;
Best Local Similarity 39.9%; Pred. No. 3.2e-25;
Matches 139; Conservative 7; Mismatches 123; Indels 79; Gaps 13;

QY 44 GYGYGLGSGAGRGGAGAAAAAGAGGGYGLGSGAGRGGLGGGAGAAAAAGGV 103
Db 2 GGGGGGSDGGGGD-----GGGSDGGGGSDGGGGSDG-----GGG 42

QY 104 GGGGLGGGAGGAGAAAAAGAGGGYGLGSGAGRGGLGGGAGAAAAAGAGAG 163
Db 43 GDGGGGSDGGGGG--SDGGGGSDGGGGGSDG---GGSGSDGG-----GGSDG 91

QY 164 GYGLGSGAGRGGLGGGAGAAAAAGAGAGGGYGLGSGAGGGYGLGSGAGRG 223
Db 92 GGGGGGSDGGGGSDGGG-----GGGSDGSDGGSDGGSDGGGGSDG-- 140

QY 224 GLGGGAGAAAAAGAGAGGGYGLGSGAGAAAAAGAG--OQGYGLGSGAGRG 280
Db 141 --GGDG-----DGGGGSDGGYGGSDGGGGGSDGGGGGSDGGSDGGG 192

QY 281 AAAAAAGAGGGYGLGGGAGGGYGLGSGAGRGGLGGGAGAAAAAGAGAGG 340
Db 193 -----GGGSDGGSDGGGGSDGGGGSDGGG---CGSDGGSDGGG 238

QY 341 GQAGAAAAAGAGAGGGYGLGGGAGGAGAAAAAGAGAGGGYGGY 388
Db 239 GDGGG--SHGGGGSDGGGGGGDAGGAD-----HDGGYG 273

RESULT 14
US-10-096-961-1
; Sequence 1, Application US/10096961
; Patent No. US20020155572A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; FILE REFERENCE: CL000849DIV
; CURRENT APPLICATION NUMBER: US/10/096,961
; CURRENT FILING DATE: 2002-03-14

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2190	81.1	718	2	A36068	major ampullate fi
2	1148	42.5	627	2	A44112	spidroin 2, dragli
3	968.5	35.9	2639	2	T31328	fibroin - Chinese
4	935	34.6	1901	2	F70806	hypothetical glyci
5	927	34.3	1079	2	B70807	hypothetical glyci
6	912.5	33.8	1489	2	D70807	hypothetical glyci
7	892	33.0	1381	2	E70806	hypothetical glyci
8	880	32.6	767	2	E70895	hypothetical glyci
9	868.5	32.2	1660	2	A70869	hypothetical glyci
10	866.5	32.1	465	1	S01820	glycine-rich cell
11	858	31.8	853	2	A70896	hypothetical glyci
12	857.5	31.8	882	2	B70812	hypothetical glyci
13	850	31.5	749	2	A70812	hypothetical glyci
14	847	31.4	1329	2	E70917	hypothetical glyci
15	844	31.3	778	2	F70963	hypothetical glyci
16	843.5	31.2	714	2	A70807	hypothetical glyci
17	839.5	31.1	914	2	H70987	hypothetical glyci
18	836	31.0	1538	2	H70846	hypothetical glyci
19	833	30.9	923	2	E70820	hypothetical glyci
20	831	30.8	1306	2	A70934	hypothetical glyci
21	828.5	30.7	641	1	Q0BE31	nuclear antigen EB
22	827.5	30.6	741	2	G70917	hypothetical glyci
23	820.5	30.4	484	2	G70846	hypothetical glyci
24	819.5	30.4	801	2	F70824	hypothetical glyci
25	818	30.3	667	2	A70893	hypothetical glyci
26	813	30.1	1011	2	F70620	hypothetical glyci
27	812.5	30.1	783	2	E70824	hypothetical glyci
28	810	30.0	731	2	C70974	hypothetical glyci
29	807	29.9	937	2	D70835	hypothetical glyci

```
QY 337 -----GGLGGGAGAAAAAGAGAGGGLGGGAG 365
Db 498 AAAGAGGEGIRGQAGAGGGYGLGSGSGRGLGGGAGAGAAAAAGAGAGGGLGGGAG 557
QY 366 QCAGAAAAAAGAGGVRGGYGGGLGSGAGRGAGAAAAAGAGGAGGGTGGGGG 425
Db 558 QGAG-----AAAAAGGVRGGYGGGLGSGAGRGAGAAAAAGAGGAGGGTGGGGG 613
QY 426 GAGGLGGGAGAAAAAGAGGGYGGVGGGASAAASRLSPQASSRVSSAVSNLVASG 485
Db 614 GRGGLGGGAGAAAAAGAGGGYGGVGGGASAAASRLSPQASSRVSSAVSNLVATG 673
QY 486 PTNSAALSTISNVVSIQAS 506
Db 674 PTNSAALSTISNVVSIQAS 694

RESULT 2
A44112
spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)
N:Alternate names: silk fibroin, dragline
C:Species: Nephila clavipes
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998
C:Accession: A44112; S27824
R:Hinman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A:Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes
A:Reference number: A44112; MUID:92406876; PMID:1527052
A:Accession: A44112
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-627 <HIN>
A:Cross-references: GB:M92913; NID:g159713; PID:g159714
A:Note: sequence extracted from NCBI backbone (NCBIP:113893)
R:Hinman, M.B.; Lewis, R.V.
submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila cla
A:Reference number: S27824
A:Accession: S27824
A:Molecule type: mRNA
A:Residues: 19-627 <H12>
A:Cross-references: ENBL:M92913

Query Match 42.5%; Score 1148; DB 2; Length 627;
Best Local Similarity 51.1%; Pred. NO. 1.3e-55;
Matches 314; Conservative 25; Mismatches 139; Indels 136; Gaps 30;

QY 17 GYG-----GLGGGAGQ-----GAG-AAAAAAGAGGGY-----GLGSGAGR--- 56
Db 3 GYGPGQQPGGYGPGQQPGSPGPGSAAAAAAGPGYGPQQPGGYGPGQQPGRYG 62
QY 57 -----GGGAGAAAAAGAGGGYGGGLGSGAGRGGLG-GQ-----GAGAAAAAGGV---- 103
Db 63 PQGQPSGPGSAAAAAGSGGPGGYGPRQGPGGYGQGGPGSGPGSAAAAASAE 122
QY 104 -GGGGLGGGAGGAGAAAAAGAGGAGGGY-----GLGSGAGRGGGGGGQ--AGAAAAA 157
Db 123 SGQQPGGYGPGQ-----GQPGGYGPGQPGGYGPGQPGSGPGSAAAAA 171
QY 158 GGAGQ-----GGYG-----GLGSGAGRGGLGGGAGAAAAAGAGGAGGGYGGGLGGGAGG 210
Db 172 SGCPQQGPGGYGPGQQPGGYGPGQQPGSPGPGSAAAAAAGSGPGGYGPGQQPG 231
QY 211 GYGGLGSGAGRGGLGGGAGAAAAAGAGGGYGGGLGGGAG-----AAAAAGGAGQ 261
Db 232 GY-----GPGQGLSGPGSAAAAAAGPGGPGGYGPGQPGSGPGSAAAAAAGP 285
QY 262 GYG-----GLGSGAG--RGQGAG--AAAAAGAGGGYGGGAGGGYGGGLGSGAGRG 315
Db 286 GYGPGQPGGYGPGQQPGSGSAAAAAAGPGQQGLGGYGPQQGPGG---YGPQQ 342
QY 316 GLGGGAGAAAAAAGAGGAGGGGLGGGAG-----AAAAAG-----GAGGG 358
```

```
Db 343 GREGYGPGSASAAAAAGPGQPGGYGPGQQPGSGPSASAAAAAAGPGYGPQQG 402
QY 359 LGGQCGAG-----GAGAAAAAAGGVRGGYGGGLGSGAG-----RGGQAG-AAAA 407
Db 403 PGGYAPGQPGSGPSASAAAAAAG--PGYGP-GQGGPGYAPGQPGSGPSAAAA 459
QY 408 AAGGAGGGTG-----GLGGQGY-----GAGGLG-----GQAGAAAAAGAGGG 447
Db 460 AAAAAAGPGYGPQPGGPAASAAAGPGYGPQPGAGYGPASAAAGAGSAG 519
QY 448 YGVCSGASASAAASRLSSPQASSRVSSAVSNLVASPTNSAALSTISNVVSIQASN 507
Db 520 YG-----PGQASAAASRLASPDGARGVASVSNLVSSGPTSSAALSVISNAVSIQASN 575
QY 508 PGLSGCDVLIQALL 521
Db 576 PGLSGCDVLIQALL 589

RESULT 3
T31328
fibroin - Chinese oak silkmoth
C:Species: Antherea pernyi (Chinese oak silkmoth)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, A
A:Reference number: 20995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: ENBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C:Genetics:
A:Introns: 14/3

Query Match 35.9%; Score 968.5; DB 2; Length 2639;
Best Local Similarity 42.6%; Pred. No. 2.2e-45;
Matches 271; Conservative 24; Mismatches 177; Indels 164; Gaps 24;

QY 20 GLGGGAGGQ-----AGAAAAAGAGGAGGGYGGGLGSGAGRGGGGAGAAAAAGAGG 73
Db 161 GAGGYYGGYSDSAAAAAAGAGAGAGGSGGSGYGGYSDSAAAAAAGAGAGAGGSGG 220
QY 74 GYGGLGSGAGRGGLGGGAGAAA---AGGVGGQ---GLGGGAGGAGAGAAAA--- 124
Db 221 GYGYYGGYSDSAAAAAAGAGAGAGGSGGSGYGGYSDSAAAAAAGAGAGAGAG 280
QY 125 ---GGAGGGYGGGLGSGQ-----GAGRG---GSGGAGAGAAAA--- 156
Db 281 AGSAGGSGGYGGYSDSAAAAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 340
QY 157 -----AGGAGGGYGGGLGSGAG-----RGG-----LGGGAGAAAAAAGGA 194
Db 341 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 400
QY 195 GGGYGGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 229
Db 401 GAGGSG--GGYGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 458
QY 230 ---AGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 278
Db 459 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 516
QY 279 A-GAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 324
Db 517 SDSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576
QY 325 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361
Db 577 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 636
```

QY 362 QGAGCAGAAAAAAGGVR-QGGYGLGSGAGRGCGGAGAAAAAAGGAGCGGTGGL 420
Db 637 YGSDSAAAAAAGAAAGAGSGGSGYGGYSDSA---AAAAAAGAGGAGCGVG-- 691
QY 421 GGGGAGGGLGGGAGAAAV-----GAGGGYGGVGSAGASAAASRLSSPQASSR 473
Db 692 GGYGWDGGYGSASAAAAAAGAGSGAGRGDGGYGSSSAAAAAASARRAG 751
QY 474 VSSAVNLVSGPTNSAALSSSTISNVVSIQIGASNP 509
Db 752 HDSAGSAAAAAAGAAAS-----GAGGSG 778
RESULT 4
F70806
hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70806
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1501 <COL>
A: Cross-references: GB:AL022022; GB:AL123456; NID: g3261554; PIDN: CAA17745.1; PID: g292444
A: Experimental source: strain H37Rv
C: Genetics:
C: Gene: RV3508
C: Superfamily: collagen alpha 1(RV) chain
Query Match 34.6%; Score 935; DB 2; Length 1901;
Best Local Similarity 37.0%; Pred. No. 1.le-43;
Matches 264; Conservative 24; Mismatches 220; Indels 206; Gaps 28;
QY 5 TGGQ--QMGRIRIRYGGGLGGGAGAGAAAAA---AAGGAG-QGGYGLGSGAGRG 58
Db 663 TCGAGSVGNAGIGTGTTGGVGGAGGAGAAAAAGSSATGGAGFAGGAGGAGGNSGV 722
QY 59 QGAGAAAAAGGAGGCGYGLGSGQAGR-----GGLGGGAGAAAAAGGVGGGLGGO 111
Db 723 GGTNGSGGAGGAGGKGTGGAGSGADNPTGAGFAGGAGGTG-GAAGAGGAGGATGTGCT 781
QY 112 GAGGAGAAAAAGGAGGCGYGLGSGAG-----RGGSGGOGAGAAAAAGG----- 159
Db 782 GGVVGTAGSAGTGGAGRGDGGDASGLGLGSLGDFDGGGGAGGAGGAGG 841
QY 160 -AGQGGYGLGSGAG-----RGLGGO-GAGAAA-----AAAAAGAGG 197
Db 842 AGNGGDDGATGAGLNDNGVGGDGGAGGAGNAGNYCLTAKAGDGGAGNGNG 901
QY 198 YGVLGGGQ-----AGQGGYGLGSGQ-----AGRGGLGGO-GAGAAA 235
Db 902 GAGGAGGAGDNNFNGGQGGAGGCGGGLGASATTSINANGAGGNGGTGKGAGGAGCT 961
QY 236 AGGAGGGLGGGA-----GAAAAAGGAGGCGYGLGSGQA-----G 273
Db 962 LGVGGSGGTGGDGDAGSGGGGFGAAGKAGGGNGRGDGGDASGLGLSLGDFDG 1021
QY 274 RGGQGAAGAAAAAGGA-GGGYGGGAGGCGYGLGSGAG-RGGLGGO-GAGAAA---- 326
Db 1022 QGGQGAGSAGAGGTNGAGGAGNG-GDGGDGTAGAGLNDNGVGGDGGAGGAGNGG 1080
QY 327 -----AAAAAGCA-----GQGLGGGQ-----AGAAAA-- 350
Db 1081 NAGVGLTAKAGDGAAGNGGAGGAGGAGDNNFNGGQGGAGGCGGGLGASATTSIN 1140

QY 351 -----AGAGCGGGLGGQ-----ACGAGAGAAAAAAGGVRQGG 386
Db 1141 ANGAGNGGTGCKGAGGAGTGLVGSCTGDDGGDGGGCGGFGAAGKAGGCGNG 1200
QY 387 YGGLGSGQA-----GRGGGAGAAAAAGG-----ACGGTGGGLGGGCV----- 425
Db 1201 VGGDGGEGASGLGLSLGSPDGGGGGGAGGAGGAGGAGGTGGAGGAGPATLI 1360
QY 426 -----GAGGLGGGAGAAAAAVGA-----GQGGYGGVGGSCASASA 460
Db 1261 GGPDDGGGGGIGDGGNAGFCAGVPGDGGDGNAGFCAGVPGDGGIGGTGAGCAGGA 1320
QY 461 AASRLSSPQASSRVSNAVNLVSGPTNSAALSSSTISNVVSIQIGASNPGLSGCD 514
Db 1321 GADGDPIDGGGAGGCGGKGKGLNSTGLASASGDDGNGGAGGAGNGGD 1374
RESULT 5
B70807
hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70807
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1079 <COL>
A: Cross-references: GB:AL022022; GB:AL123456; NID: g3261554; PIDN: CAA17749.1; PID: g292422
A: Experimental source: strain H37Rv
C: Genetics:
C: Gene: RV3512
C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
Query Match 34.3%; Score 927; DB 2; Length 1079;
Best Local Similarity 44.3%; Pred. No. 1.9e-43;
Matches 235; Conservative 15; Mismatches 202; Indels 78; Gaps 22;
QY 7 GQMGIRIRYGGGLGGGAGGAGAAAAAAGGAGCGYGLGSGAGRGCGGAGAAAA 66
Db 212 GATPGANGIAGNGDGDGA---AGAVGISGATGAGDGGHGTGAAGNGGTGGAGSGI 268
QY 67 AAGGAGCGYGLGSGQA-----GRGGLGGQ--AGAAAAAGG-----VGQ 105
Db 269 DGVGGTGTGGNGNGATGGAGDAGGSGNSGNGIGGKGNAGAGGAGGNGGTGCA 328
QY 106 GGLGGGAGGAGAAAAA---GGAGCGYGLGSGAGRGSGGCGGAGAAAAAGAG-- 161
Db 329 NGTGGDGGNGRAGATAGSNGAGGTGSAGNGGTG-GRGSGGAGGDTGGVGGKGN 387
QY 162 -----QGGYGLGSGQ-----AGRGGLGGQ-----GQGG 198
Db 388 GADGEYGGAGGAGGSPNTSPGNGGQGGGSGGAGGAGGAGGANGTAGNGGCGG 447
QY 199 YGGLGGQA-----GQGGYGLGSGAGRGGLGSGGAGAAAAAGGAGCGGGLGGGAGA 251
Db 448 AGGTGGAGAASATNGSGGAGGTGGD-GSGGAGTG-CAGCTGGAAGDGGCGG-CA 504
QY 252 AAAAAAGGAGGCGYGLGSGAGRGCGAGAAAAA---AGGAGCGYGGCGAGCGYGLGGS 309
Db 505 GGGAGCGGAGGAGGTGGNGNITGCTAGTAGNAGNGRAGKAGGAGGCGGTGGTGGCGG 564
QY 310 QGAGRGGLGGQ-----AGAAAAAAGG-----AGCGGLGQCA--CAAAAAAGAGCGGL 359
Db 565 AG-GDGGAGGTGGDRTVGGTVPAGSCGCGNAGGCGGAGCGGCGGADGGCGDAGTGGN 623
QY 360 GGGGAGCGAGAAAAAAGGVRQGGYGLGSGQAGRGGO-GAGAAAAAAGGAGGCGGTG 418

Db 1125 GAGGCGAGGCGGIGTGGAGGAGGCGDPCGAGGCGGLGDSGNGTSAANGVDSKHKGP 1184
Qy 426 ---GAGGLGGCGAGAAAAGV-GAGQGGYGG-----VSGSASASAAA 462
Db 1185 LTGGDGGVGNCAKAAAAGDGGCGDGNAGLFGDGGAGGDACTAAEA 1235

RESULT 10

S01820
glycine-rich cell wall protein 1.8 precursor - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 30-Sep-1989 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: S01820
R:Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A:Title: Glycine-rich cell wall proteins in bean: gene structure and association of the
A:Reference number: S01820; MUID:89091109; PMID:3208742
A:Accession: S01820
A:Molecule type: DNA
A:Residues: 1-465 <REL>
A:CROSS-references: EMBL:X13596; NID:g21002; PIDN:CAA31932.1; PID:g21003
C:Comment: This protein is enriched in the cell wall fraction of young hypocotyls and ov
tyls.
C:Comment: Much of the sequence consists of tandemly repeated 22-residue segments with t
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8
C:Keywords: cell wall; structural protein; tandem repeat
F:1-30/Domaln: signal sequence #status predicted <SIG>
F:31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

Query Match 32.1%; Score 866.5; DB 1; Length 465;
Best Local Similarity 47.8%; Pred. No. 1.8e-40;
Matches 218; Conservative 8; Mismatches 177; Indels 53; Gaps 19;

Qy 23 GCGAGCGAGAAAAAGGAGCGGCGGLGSGQAGRGCGGAGAAAAAGAGCGGCGGLGSGQ 82
Db 33 GYGLHGTGCGYGAGAGSYGGCGGGGGGGGGYAGEHVYVGGSGGGGGGVTGGGQ 92
Qy 83 GAGRGGLGSGQAGAAAAGGCGGGLGSGQAGCGAGGAGAAAAAGGAGCGGCGGLGSGQAGR 142
Db 93 GAGYGGGGGGGGGVAYGGGGE--RGGYGGGCG-----GGAG-GGYGAGGEHGIY 141
Qy 143 GSGSGCGAGAAAA--RAGGAGCGGCGGGLGSGQAGRG-----LGGQAGAAAAAGAG 193
Db 142 GGGGGGAGGGGGYAGGAGGGGCGTGGAGGGGGGGGDDHGGGCGGAGGAGGGGCGG 201
Qy 194 AGCGGCGGLGSGQAGG--GGYGGGCGGAGCGGGLGSGQAGAAAAAGAGCGGGLGSGQAGAA 252
Db 202 GGEHGGGGGGGGGAGGCGGCGGAGCGGCGGAG--GGCGGAGCGGCGAGGEGH--GGAGGGGQ 258
Qy 253 AAAGAGCGGCGGGLGSGQAGRGCGGAGAAAAAGAGCGGCGGCGGCGGCGGLGSGQAG 312
Db 259 GGAGGCGAGGEGH--GGAGGGCGG--GAGGCGGAGGCGGAGGGGGGGGAGGGYAGGEGH 316
Qy 313 GRGGLGGCGAGAAAAAGGAGCGGCGGCGGAGAAAAAGGAGCGGGL--GGCGAGCGAG 370
Db 317 GGGG--GGCGGAGGGGAAVAGGCGGCGGCGG-----DGGGCGTGGGCGGCGGCGG- 370
Qy 371 AAAAAAAGGVRQGGCGGGLGSGQAGRGCGGAGAAAAAG--CAGCGGTGGL-----420
Db 371 --AGGCGTGGGEGGCGG--GGCGGG--GGYAGGCGHGAAGYGGGEGGGCGGCGGCGGAG 425
Qy 421 --GGGCGAGGCGGCGAGAAAAAGGCGGCGGCGG 454
Db 426 HGGGCGGAGGGGGYAGGAGH--GGYGGGGGIGGG 459

RESULT 11

A70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COL>
A:CROSS-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17207.1; PID:el25
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match 31.8%; Score 858; DB 2; Length 853;
Best Local Similarity 42.8%; Pred. No. 8.4e-40;
Matches 232; Conservative 18; Mismatches 184; Indels 108; Gaps 26;

Qy 7 GQMGRIIRIRGYGGLGGGAGGAGAGAAAAAGAGAGCGGCGGLGSGQAG--GRGGGAGAG 63
Db 292 GSGGSHALLWAGAGAGGCGGCTGAGGCTAGACGNGGAGCGGCTGGLFNGGAGGHHG 351
Qy 64 AAA-----AAGGAGCGGCGGGLGSGQAGRGGLGCG-----GAGAAAAAGGCGGGL 108
Db 352 AAAGNLAAGNCGVSSSGGGGAGGTGGAGGCG--GAGGAGNARLWGVGGAGG--GDGGA 408
Qy 109 GCGAGCGAGAAAAAGGA-----GCGGCGGLGSGQ-----ACRGGSGGCGA----- 150
Db 409 CGAGCGGCGSLSGNAGGAGCGDSRGCTGAGGCGGAGGLVGTGCHGGDGGAGGAAYKG 468
Qy 151 --GAAAAAGGAGCGGCGGLGSGQAGRGGLGSGAGAAAA-----AAAA 191
Db 469 GGGGAAAGTGTAGAGGCGAG-----GSGGSGDGGGGAAGPAGWLFPGDGGAGGNGGAAAA 524
Qy 192 CGAG--QGGYGGCGGAGCGGCGGGLGSGQAG-----GRGGLGGCGAGAAAAAGGA----- 239
Db 525 GGAGGAGGCGGGGNGG--GNGGNGGNGGATGWLNGGAGGAGGATAGAGAGANGVS 583
Qy 240 -----GCGGLGCG--GAGAAAAAGAGCGGCGGLG--SQAG--RGCGAGAAAAAGG 288
Db 584 STNGGCTGGNGGCTGCGGCGAGNAGLLGVGAGGCGHAGGAGGCGGAGGCTGCTSSDG 643
Qy 289 AGCGGCGGCGAGCGGCGGGLGSGQAGRGGLGCG--GAGAAAAAAGAGCGGCGGLGSGAGAA 347
Db 644 A--GGDGGCG--CNGAGGTGGLTFAGGNGGCGGSGGAADICGNGAGMCG--GTGNGNG 699
Qy 348 AAAAGCGCGGLGCGGAGCGA-----GAAAAAAGAGGVRQGGY--GLGSGQ-- 394
Db 700 GSGGCGAGSGDGGGAGGNGAMLFNGGAGGGGCGGNGAGGGLGSGGSLPLNGSGGD 759
Qy 395 ACRGGCGA--GAAAAAGAGCGGCTGGLGCGG--GAGGLGCGGAGAAAAAGGCGGCGG 450
Db 760 GCGGNGAPGCVLYLVNCGGAGCGGCGGCTGGGATGCGAGGCGGCGGCGGCGGCGG 819
Qy 451 VG 452
Db 820 AG 821

RESULT 12

B70812
hypothetical glycine-rich protein Rv0834c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70812
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

C:Accession: E70917
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: E70917
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1329 <COL>
A: Cross-references: GB:295844; GB:AL123456; NID:3250713; PIDN: CAB09271.1; PID: g2131046
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: RV1450c
C: Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 31.4%; Score 847; DB 2; Length 1329;
Best Local Similarity 38.9%; Pred. No. 4.8e-39;
Matches 243; Conservative 16; Mismatches 196; Indels 170; Gaps 27;

```
QY 5 TGGQQMGRIRIRYGGGL---GGGAGQGAGAAAAAAGAGAGGGYGGGLGSGQAGRGGQGA 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 SGGKAGGNGGAGGAGGLVNGGAGGAGNGACPCAPSSGDPNGGGGGAGAG--GKGGDG-- 466

QY 62 GAAAAAAGGAGGGYGGGLGSGQA--GRGLGGGAGAAAAAGYGGGGLGGGQAGGAG-- 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 GAQAGDGGAGGAGGKGGNGGATGATGLNGLGAGADGTDG--GKGGNGAGGGGAGGQ 524

QY 119 -----AAAAAGAGGCGYGGGLGSGQA--GRGGGGGQAGAAAAAGAGGQ 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 525 GKKALAAHQDQSMGAGGAGGNGGAGGCGDGGNGAKGTFDNGDGVGNGGNGSRGIG 584

QY 164 YGVGGLGSGQA-----GRGLGGGAGAAAAAAGGA--GOGYGGGLGSGQA 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 585 GAGGICAGSTAGADCARATPTSCGGGTGGNGANATVAGGAGGAGGKGGNGGLVNG-- 643

QY 208 GGGYGGGLGSGQ-----AGRGGLGGQ--GAGAAAAAGG-----AGOGGLGGQGA 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 644 GAGGKGGDMAGYAGSSPTTAGESGTSQNGGAGGAGGAGRGGRDFFGDDGTTGGAGNGA 703

QY 250 GAAAAAAGGA--GOGYGGGLGSGAGRGGQGA-----GAAAA 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 704 NGANATTPKAGKGDDGHGGPFAQG--GNGGQGGPGLAGNLFQNGITQYGGSGKGGCAGG 762

QY 285 AAG-----CAGOGGYGGGQ-----AGOGYGGGLGS--QGA-----GRG 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 763 LAGDGGNGANGNFAFGDNGGHHGNGCNPCAGCQGGSGGAGSTPGAKGAHGFTPTSGGDG 822

QY 316 GLGGGQ-----AGAAAAAAGGAGGGGLGGQGA----- 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 823 GDGGNGNSQVYGGNGDGGNGGNGSGSAGCTGGNGRRGGDAGFGMSANATNPENGPNGN 882

QY 345 ---GAAAAAAGGAG--QGGILGGGAGGAGCAAAAAAAGGVR-----QGGYGGGLG 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 883 PGNGGAGGAGGAGLNGNGGAGGNGGLGGFFGNGAAGANGVAVGAPQPGGAGGHHGAG 942

QY 392 SQG--AGRGQ-----GAGAAAAAAGGAGGCTGGILGGGQVG---AGILGGGAGAAAA 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 943 GNGGAGGNGCQGVVSDGACAGACGCGDGPAGDANGNGGCGQAGAPAGGGGGRGDDGNA 1002

QY 441 VGAGOGGYGGVSGASAAASRL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1003 GNAGAGPGGTGTAGKAGPAGSIL 1027
```

RESULT 15
F70963
hypoetical glycine-rich protein RV2634c - Mycobacterium tuberculosis (strain H37RV)
C: Species: Mycobacterium tuberculosis
C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C: Accession: F70963

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70963
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-778 <COL>
A: Cross-references: GB:280225; GB:AL123456; NID:3242265; PIDN: CAB02341.1; PID: e26639
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: RV2634c
C: Superfamily: unassigned collagens

Query Match 31.3%; Score 844; DB 2; Length 778;
Best Local Similarity 38.9%; Pred. No. 4.5e-39;
Matches 225; Conservative 24; Mismatches 204; Indels 126; Gaps 25;

```
QY 2 ASMTGGQ-----QMGRIRIRYGGGLGSGQAGQ--AGAAAAAAGGA--GOGYGGGLGSGQ 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 ATLVGTTGGVGGATGLGSGFGGAGGAAAGVCTTGGVGGSGVGVFGNGFGGAGGLG 266

QY 54 AGRGGGAGAAAAAAGGAGGQGGYGGGLGSGQA-----GRGLGG--QGAGAAAAAGG-- 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 AAGGVCGAASYGTGTGGGGVGGDAPGDDGAGPLLIINGVGGVGLGAGAGNNGGAGGM 326

QY 103 -VGQGGILGSGQ-----AGQGAGAAAAAAGGA--GOGYGGGLGSGQAG----- 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 LLGDDGAGGQGGPAPVAGVLGGMPGAGGNGGNANWFSGGAGGCGGTGLAGTNGVNPGSIA 386

QY 142 -----RGSSGGGAGAAAAAAGGAGGCGYGGGLGSGQAGRGGLGQ--GA 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 NPWTGANGTDSNGNQTTGGNGGPGPAGVGGEAGV--GGQGGGLGSLDNDGTGCKGGA 444

QY 184 GAAAAAAGCAGGCGYGGGLG-----CQAGCQGGYGGGLGSGQ-----AGRGGLGCGQAGA 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 GGTAGTDDGAGGAGGAGGIGETDGSAGGYATGEGGDDGATGGVGGVGGAGKGGGQHNW 504

QY 233 AAAAGGAGCGGLGGQAGAAAAAAGCAGGCGYGGGLGSGQ-----AGRGGQAG 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 505 GVGDAFGDGGIGCGDNGALGAAGNGGTCGAGNGGRRGMLITNGCAGCAGGTGCTGGG 564

QY 281 AAAAAAGGAGCGYGGQ-----AGOGYGGGLGSGQ-----AGRGGLGCGAGAAA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 565 GAAGFAGV--GGAGGGLTDGAGTAEGGTGGILGGVGGTGTGGMGSGGVGGNGGAAGS 622

QY 327 AAAAGGAGCGGLGGQAGAAAAAAGCAGGCGGLGGQAGCGAGAAAAAAGGAGVRRQG 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 LITLGGGGAGGVGGTG-----GIGGIGGAGNGGAGGAGTTTGGGATIGGGGTGG 674

QY 387 YGGLGSGQ--AGRGQ--GAGAA-----AAAAGAGCGGT-----GGLGGGAGGAGLGGQ 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 675 VGGAGGTGTGGAGGTTGGSGGAGLIGWACAGGTGAGGTGGGQGLGGQ--GNGGNGT 733

QY 434 GA----GAAAAACAGGCGYGGVSGASAAASRLSSP 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 GATGGGQGGDFALG--GNGGAGGAGGSPGSGSGTQGNMGPP 771
```

Search completed: December 18, 2002, 16:22:22
Job time : 21.6009 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 8.10437 Seconds
(without alignments)
2702.183 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700
Sequence: 1 MASMTGGQMGRRIRINGGG.....GLSGCDVLIQALLGHHHHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2273	84.2	747	1 SPDI_NEPCL	P19837 nephila cla
2	1148	42.5	627	1 SPDI_NEPCL	P46804 nephila cla
3	952.5	35.3	5263	1 FBOH_BOMMO	P05790 bombyx mori
4	935	34.6	1901	1 Y208_MYCTU	O53553 mycobacteri
5	866.5	32.1	465	1 GRP2_PHAVU	P10496 phaseolus v
6	844	31.3	778	1 YQ34_MYCTU	P71933 mycobacteri
7	839.5	31.1	914	1 WA22_MYCTU	O06794 mycobacteri
8	828.5	30.7	641	1 EBN1_EBV	P03211 epstein-bar
9	819.5	30.4	801	1 Y747_MYCTU	O53810 mycobacteri
10	807	29.9	957	1 Y278_MYCTU	P56877 mycobacteri
11	782.5	29.0	603	1 YD25_MYCTU	O10637 mycobacteri
12	736	27.3	384	1 GRP1_PETHY	P09789 petunia hyb
13	729.5	27.0	860	1 ELS_MOUSE	P54320 mus musculu
14	714.5	26.5	338	1 LRP_ARATH	P27483 arabidopsis
15	697.5	25.8	481	1 LOR1_MOUSE	P18165 mus musculu
16	684	25.3	543	1 YP91_MYCTU	O50630 mycobacteri
17	680.5	25.2	864	1 ELS_RAT	O99372 rattus norv
18	671.5	24.9	498	1 Y118_MYCTU	O50615 mycobacteri
19	612	22.7	463	1 Y468_MYCTU	O53416 mycobacteri
20	609.5	22.6	515	1 Y140_MYCTU	O50594 mycobacteri
21	602	22.3	491	1 YK98_MYCTU	O10707 mycobacteri
22	595	22.0	747	1 ELS_BOVIN	P04985 bos taurus
23	593.5	22.0	750	1 ELS_CHICK	P07916 gallus gall
24	536.5	19.9	730	1 ELS_HUMAN	P15502 homo sapien
25	526.5	19.5	252	1 GRP1_PHAVU	P10495 phaseolus v
26	516	19.1	1733	1 VNUA_PRVKA	P33485 pseudorabie
27	512.5	19.0	672	1 PHX5_MOUSE	P08399 mus musculu
28	501.5	18.6	1356	1 CA21_ONCMY	O93484 oncorhynch
29	471.5	17.5	1355	1 CA21_RANCA	O42350 rana catesb
30	470	17.4	2038	1 FSH_DRONE	P13709 drosophila
31	455.5	16.9	966	1 FIB1_PETMA	P02674 petromyzon
32	454	16.8	316	1 LORI_HUMAN	P23490 homo sapien
33	439	16.3	671	1 CA11_RAT	P02454 rattus norv

ALIGNMENTS

RESULT 1

ID	SPDI_NEPCL	STANDARD;	PRT;	747 AA.
AC	P19837;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidea; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-1- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; M37137; AAA29380.1; -			
DR	EMBL; U03848; AAB60212.1; -			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	1	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	655	1.
FT	REPEAT	1	25	2.
FT	REPEAT	26	38	3.
FT	REPEAT	39	66	4.
FT	REPEAT	67	96	5.
FT	REPEAT	97	130	6.
FT	REPEAT	131	158	

34	434	16.1	1027	1	CAFF_RIPPA	P30754 riftia pach
35	432	16.0	1156	1	GLH4_CAEEL	O76743 caenorhabdl
36	429.5	15.9	1453	1	CA11_CHICK	P02457 gallus gall
37	428.5	15.9	1466	1	CA13_HUMAN	P02461 homo sapien
38	428	15.9	1464	1	CA11_MOUSE	P11087 mus musculu
39	427	15.8	1453	1	CA12_MOUSE	P28481 mus musculu
40	422	15.6	1459	1	CA11_CANFA	O9x917 canis famil
41	420	15.6	1460	1	CA11_CANFA	Q01119 mus musculu
42	419	15.5	1372	1	CA21_MOUSE	O46392 canis famil
43	415.5	15.4	1366	1	CA21_MOUSE	P02458 homo sapien
44	415	15.4	1418	1	CA12_HUMAN	P08123 homo sapien
45	412.5	15.3	1366	1	CA21_HUMAN	

```
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 662
FT CONFLICT 672 672
FT CONFLICT 695 747
SQ SEQUENCE 747 AA: 60528 MW: 850E44B0D649E012 CRC64:
Query Match
Best Local Similarity 84.2%; Score 2273; DB 1; Length 747;
Matches 494; Conservative 4; Mismatches 10; Indels 188; Gaps 15;
QY 6 GQQQMRIRIRYGGILGGQAGCGAGAAAAAGAGGGYGGILGSGQAGRGQGGAGAAA 65
DB 22 GQGGAGQ---GGYGGILGGQAGCGAGAAAAAGAGGGYGGILGSGQAGRGQGGAGAAA 78
QY 66 AAAGGAGGGYGGILGSGQAGRGGLGGQAG--AAAGAGGGYGGILGG--QGAGGCA-CA 119
DB 79 AAAGGAGGGYGGILGSGQAGRGGLGGQAGAAAAAGAGGGYGGILGSGQAGRGQGGAG 138
QY 120 AAAAGGAGGGYGGILGSGQAGRGSGGGQGGAGAAAAAG--GAGGGYGG 166
DB 139 AAAAGGAGGGYGGILGSGQAGRGGLGGQAGCAAAAAAGAGGGYGGILGGQAGGGYGG 198
QY 167 GLGSGQAGRGGLGGQAGGAG----- 185
DB 199 GLGSGQAGRGGLGGQAGCAAAAAAGAGGGYGGILGGQAGCGAGAGAGAGGGYGG 258
QY 186 -----AAAAAGGAGGGYGGILGGQAGGGYGGILGGQAGCGAGAGAGGGYGG 232
DB 259 GSGAGRGGGEGAGNAAAAGAGGGYGGILGGQAGGGYGGILGGQAGRGGLGGQAGG- 317
QY 233 AAAAGGAGGGYGG---GQAGAAAAAGAGGGYGGILGSGQAGR---GGQAGCA-AAA 284
DB 318 AAAAGGAGGGYGGQAGGAGCAAAAAAGAGGGYGGILGSGQAGRGGLGGQAGGAVAAA 377
QY 285 AAGGAGGGYGG---QGAGQ-----GAGGGYGGILGSGQAGRGGLGGQAGCA 324
DB 378 AAGGAGGGYGGILGSGQAGRGCGGAGAAAAAGAGGGYGGILGSGQAGRGGLGGQAGCA 437
QY 325 AA-----AAAAAGGAGQ----- 336
DB 438 AAAAGGAGGGYGGILGSGQAGRGCGGAGAAAAAGAGGGYGGILGSGQAGRGCGGAGAAA 497
QY 337 -----GGLGGQAGAGAAAAAGAGGGYGGILGSGQAGRGCGGAGAAA 365
DB 498 AAAGVAGQBGIRIQGAGGGYGGILGSGQAGRGGLGGQAGAGAAAAAGAGGGYGG 557
QY 366 QGAGAAAAAGAGGGYGGYGGILGSGQAGRGCGGAGAAAAAGAGGGYGGILGGQGV 425
DB 558 QGAG-----AAAAAGGGYGGYGGILGSGQAGRGCGGAGAAAAAGAGGGYGGILGGQGV 613
QY 426 GAGGLGGQAGAAAAAGAGGGYGGYGGVGSAGASAAASRLSPQASSRVSSAVSNLVASG 485
DB 614 GREGILGGQAGAAAAAGAGGGYGGYGGVGSAGASAAASRLSPQASSRVSSAVSNLVASG 673
```

```
QY 486 PTNSAALSSTISNVYSQIGASNPGLSGCDVLQALL 521
DB 674 PTNSAALSSTISNVYSQIGASNPGLSGCDVLQALL 709

RESULT 2
SPD2_NEPCL
ID SPD2_NEPCL STANDARD; PRT; 627 AA.
AC P46804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spidroin 2 (Dragline silk fibroin 2) (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92406876; PubMed=1527052;
RA Hinman M.B., Lewis R.V.;
RT "Isolation of a clone encoding a second dragline silk fibroin.
RT Nephila clavipes dragline silk is a two-protein fiber.";
RL J. Biol. Chem. 267:19320-19324(1992).
CC -!- FUNCTION: Spiders major ampullate silk possesses unique
CC characteristics of strength and elasticity. Fibroin consists of
CC pseudocrystalline regions of antiparallel beta-sheet interspersed
CC with elastic amorphous segments.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- DOMAIN: Highly repetitive protein characterized by regions of
CC polyalanine and glycine-rich repeating units.
CC -!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M92913; AAA29381.1; -.
KW Silk; Repeat.
FT NON_TER 1 530 15 APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1 36 1.
FT REPEAT 37 79 2.
FT REPEAT 80 121 3.
FT REPEAT 122 172 4.
FT REPEAT 173 213 5.
FT REPEAT 214 252 6.
FT REPEAT 253 283 7.
FT REPEAT 284 317 8.
FT REPEAT 318 359 9.
FT REPEAT 360 391 10.
FT REPEAT 392 428 11.
FT REPEAT 429 464 12.
FT REPEAT 465 488 13.
FT REPEAT 489 515 14.
FT REPEAT 516 530 15.
SQ SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;

Query Match
Best Local Similarity 42.5%; Score 1148; DB 1; Length 627;
Matches 314; Conservative 25; Mismatches 139; Indels 136; Gaps 30;
QY 17 GYG-----GLGGQAGQ-----GAG-AAAAAAGAGGGYGG-----GLGSGQAGR--- 56
DB 3 GYGPGQGGPGYGGPGQGGPGSGPSAAAAAAGAGGGYGGPGQGGPGYGGPGQGGPGYGG 62
QY 57 -----GGQAGAGAAAAAGGAGGGYGGILGSGQAGRGGLG-GQ-----GAGAAAAAGGV- 103
```


Db 666 GAAGLFANGAGGAGGTGTAGGAGGAGGAGGLYAHGTTGGPGGNGSTCAGCTGAGGP 725

QY 436 ----GAAAAGVAGC-----QGGYGGVCGSASAAAASRLSSPQASSRVSSAVNLVASGP 486

Db 726 GGLYGAGGGAGGCHGGMAGGGGCGGNAGSLTLNAGGAGGGSGGSLSGKAG---AGGA 782

QY 487 TNSAAL 492

Db 783 GGSAGL 788

RESULT 8

EBNL_EBV STANDARD; PRT; 641 AA.

AC P03211;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE EBNA-1 nuclear protein.

GN BKRF1

OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammapesvirinae; Lymphocryptovirus.

OX NCBI_TaxID=10377;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84270667; PubMed=6087149;

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuffnell P.S., Barrell B.G.

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";

RL Nature 310:207-211(1984).

RN [2]

RP SEQUENCE OF 1-26 FROM N.A.

RX MEDLINE=86259739; PubMed=3460083;

RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;

RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear

RT proteins: a probable transcriptional initiation site.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).

RN [3]

RP SUBCELLULAR LOCATION.

RX MEDLINE=90266473; PubMed=21611150;

RA Pettit L., Sample C., Kieff E.;

RT "Subnuclear localization and phosphorylation of Epstein-Barr virus

RT latent infection nuclear proteins.";

RL Virology 176:563-574(1990).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.

RX MEDLINE=96006523; PubMed=7553871;

RA Bochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,

RA Edwards A.M., Frappier L.;

RT "Crystal structure of the DNA-binding domain of the Epstein-Barr

RT virus origin-binding protein EBNA 1.";

RL Cell 83:39-46(1995).

CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE

CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIVATING FACTOR

CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.

CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT

CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED

CC WITH THE NUCLEAR MATRIX.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: V01555; CAA24816.1; -

CC EMBL: M13941; AAA45889.1; -

CC PIR: A03773; Q0BE31.

CC PIR: S33021; S33021.

DR 1VHI; 23-DEC-96.

DR TRANSFAC; T00211; -

DR InterPro: IPR004186; EBNAL.

DR Pfam: PF02905; EBNAL1; 1.

KW Nuclear protein; DNA-binding; Transcription regulation; Activator;

FT 3D-structure.

FT DOMAIN 87 352 GLY/ALA-RICH.

SQ SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

Query Match 30.7%; Score 828.5; DB 1; Length 641;

Best Local Similarity 46.0%; Pred. No. 6.4e-32;

Matches 209; Conservative 18; Mismatches 152; Indels 75; Gaps 17;

QY 40 GAGCGYGGGLGSGAGGCGGAGAGGAGGAGGCGGCGGGLGSGAGRGGLGGGAGAGAAA 99

Db 5 GPGTGPCNGLGKGDTSPEG-----SGSGPQRRGG-DNHGRCGRGRCGRGPRGA 56

QY 100 AGGVGOG-----GLGGQ--GAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140

Db 57 PGSGSGPRHRDGVRRPQKRPSCIGCKGTGHTGAGAGAGAGAGAGAGAGAGAGAGAG 116

QY 141 GRGSGGCGGAG 200

Db 117 GAGGAGGAGAG-----GGAGAGS-----GAGAGGAGAG-----GGAGAG--G 152

QY 201 GLGCGAGAGCGYGGGLGSGAGRGGLGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260

Db 153 GAGGAGAG--CGAGGAG--CAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 203

QY 261 QGCGYGGGLGSGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320

Db 204 AGGAGAG--GAG 261

QY 321 GAG 380

Db 262 GAGGAG 320

QY 381 GVRQGGYGGGLGSGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440

Db 321 G-----AGAGGGGGRGGRGSG--GRGRGGSGGRGSGGRRGRGRER----ARGGSRERA 369

QY 441 VGAGOG-GYGGVCGSAS 473

Db 370 RGRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 403

RESULT 9

Y747_MYCTU STANDARD; PRT; 801 AA.

AC O53810;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical PE-PCRS family protein RV0747 precursor.

GN RV0747 OR MT0772.5 OR MV041.21.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala E.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Horsley T., Jorgensen K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

[illegible]

```

Db 219 GLFAGAGGAGGPGFNG-GAGGAGRSFLFEVLAAGGA--GTTGLSVNG-GTGTGGT 274
QY 182 GAGAA--AAAAAGGAG-----QGQGGYGLGSGQ-----AGRG 223
Db 275 GGGGLFNSGAGGAGGFGVSGSAGNGGTGGDGIPTGNGGTGGTGTGNQLVGGEG 334
QY 224 GLGGQAGAGAAAGGAGGGLGQAGAGAAAAGGAGGQGGYGLGSGQA-----GRG 275
Db 335 GAGGAGNAGILFAGAGIGGTGTGTGLG-APDPGCTGKGVGGIGGAGALFPGGAGGTG 393
QY 276 GOGAGAAAAAAGG-AGQGGVGGQG--NGQGGYGLGSGQ-----GAGRGLGGQAGAA-AA 327
Db 394 GFGASSADQWAGIGGSGSGGAAKILIGDGGAGTGGDSVVRGANGSGGTGGTGLIDGG 453
QY 328 AAAAGAG-----QGGLGGQAGAAAAAGGAGQGGGLGGQAGAGAAAAAAGGV 382
Db 454 AGGAGGTGTEFGSVGAGGAG-GNAAGLSGAGGAGGAGGAGGAGGAGGAGGAGGAGG 512
QY 383 ROGYYGLGSGQ-AGRGQAGAGAAAAGGAGGQGGTGGGLGGQGVGA-GGLGGQAGAAAA 440
Db 513 GAGGAGGLGIAGDGGNGRG-----GKAGMVGNGSDGAGAGSVVANGVGGSGGNATLI 567
QY 441 VGAGQGGYGGVGS--CASASAAASRL---SSPOAS 471
Db 568 GNGNGGNGVGSAPKGGAGGTAGLLGNGSPGLS 603

RESULT 12
GRP1_PETHY
ID GRP1_PETHY STANDARD; PRT; 384 AA.
AC P09789;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RA Condit C.M., Meadner R.B.;
RT "A gene encoding a novel glycine-rich structural protein of petunia.";
RL Nature 323:178-181(1986).
CC -|- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -|- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -|- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC -|- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC -|- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40
AA
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC DR EMBL; X04335; CAA27866.1; -.
CC PIR; A26099; A26099.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 41 384 GLY-RICH.
FT SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

```

```

Query Match 27.3%; Score 736; DB 1; Length 384;
Best Local Similarity 46.8%; Pred. No. 7.5e-28;
Matches 197; Conservative 10; Mismatches 130; Indels 84; Gaps 22;

QY 40 GAGQGGYGGILGSGQAGRG---GQAGAAAAAGAGAGQGGYGLGSGQAGRGGLGGQAGAGA 96
Db 41 GCQFGRRCGGGGRFGRGPFGRGRGAGGFGGGAGGGAGGGLG---GGGGLGGG--- 93
QY 97 AAAAGVGQGGILGCGQAGAGAGAAAAAGAGAGQGGYGLGSGQAGRGSGGQAGAGAAAA 156
Db 94 -GAGGGGLGGGCGAGGGFG-----GGAGGAGGGLG-CGGLGGGCGGAGGGGGV 143
QY 157 AGGAGQGGYGLGSGQAGRGGLGGQAGAGAAAAAGAGAGQGGYGLGSGQAGAGGGGLG 216
Db 144 GGGAGSGGGFGAG-----GGVG--GAGA-----GGVGGGGGFG-----GGGGVG 184
QY 217 SQAGRGGLGGQAGAGAAAAAGAGAGQGGGLGGQAGAGAAAAAGAGAGQGGYGLGSGAGRG 276
Db 185 G-GSGHGGFGAGGGVGGAGGGLGGVG-----GGGGSGGGGIG-----GG 228
QY 277 QGAGAAAAAGAGAGQGGYGLGSGQAGRGGLGGQAGRGGLGGQAGAGAAAAAGAGAG 336
Db 229 SGHGGCGFAGGGVG-CGVGGAGGGGGG-GGGGGGGGLG-GSGHGGGFGAGGGVG 285
QY 337 GGLGGQAGAAAAAGAGAGQGGGLGGQAGAGAAAAAGAGAGYRQGGYGLGSGQAG 396
Db 286 GAAGGVGGGGFGGGGGGGVGGSGHGGFGAG-----GGVGGAGGGGLG----- 330
QY 397 RGGQAGAAAAAGAGAGQGGTGLGGQGVGAG-GLG-GQAGAGAAAVGAGQG--GYGGVGS 453
Db 331 -GGGGAG-----GGGGIGGGHG-GGFGVGVGIGVGVGAGAGAGVGGVGGSGGGGN 382
QY 454 G 454
Db 383 G 383

RESULT 13
ELS_MOUSE
ID ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BA1B/C; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RT Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
mouse chromosome 5 in a region of linkage conservation with human
chromosome 7.";
RL Genomics 23:125-131(1994).
CC -|- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -|- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
INTO AN EXTENSIBLE 3D NETWORK.
CC -|- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -|- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC

```

```
CC -----
DR EMBL; U08210; AAA80155.1;
DR MGD; MGI:95317; Ela.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
SQ SEQUENCE 860 AA; 71955 MW; 0C0BE5AAE1EDD7F1 CRC64;

Query Match 27.0%; Score 729.5; DB 1; Length 860;
Best Local Similarity 42.6%; Pred. No. 2.5e-27;
Matches 226; Conservative 23; Mismatches 179; Indels 103; Gaps 25;

QY 15 IRGVGGLGGGCGAGCAAAAAGAGGAGGGYGGGLG-SQAGRGG-----Q 59
Db 293 VAGAGGKAGYPTGTGTGVSQAASAAAKAYGAGGAGYLPVGGGGIPCGAIPGIGIA 352
QY 60 GAG-----AAAAAGGAGGGYGGGLG-----QGAGRGGLG-OCAGAAAAAGGVGQ 105
Db 353 GAGTAAAAAATAAATAAAYAGGLVPGPGVRLPGAGIPGVGIPGVGIPGVGPGI 412
QY 106 GG-----LGGCGGCGAGAAAAAGAGGGYGGGLG--SQAGRGSGGGGAG----- 151
Db 413 GPGGIVGPGAVSPAAAAAATAAATAAAYAGGAGGAGGAGGAGGAGGAGGAGGAG 472
QY 152 -AAAAAGGAGGGYGGGLGSGQA-----GRGGLGGGQ--AGAAAAAA 190
Db 473 PAAAAAATAAAYAGGAGGAGAGLGLVPGAVPALPGAVPAGGVPAGGAGTAAAAA 532
QY 191 AGGAGCGGAG-GLGG--QGAGGGYGGGLGSGAGRGGLGGGAGAAAAAG-GAGOGGLG 246
Db 533 AKAAKAGLPGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 587
QY 247 QGAGAAA-AAAGAGCGGGYGGGLGSGAGRGGGGAGAAAAAG-GAGGGYGGGAGGGY 304
Db 588 AGSPAAAKSAKAAKAAQYRAAAGLGGVPGFAGAGVPGFAGAGVPGFAGAGVPGF 646
QY 305 G-GLGSGAGRGGLGGGAGAAAA--AAAGAGCGGLGGGAGAAAAAGGAGCGGLG 361
Db 647 GAGAGVPGFAGAVPGSLAASAKAAQYAGAGLGGPGGLGGPG-----GLGPGGLG 698
QY 362 QG-----AGCAGAAAAAAAGAGVVRQ---GGYGGGLGSGAGRGGAGAAAAAGAG 413
Db 599 AGVPCRVAGAAPAAAAAATAAATAAAYAGLGGAGLGGAGLGGAGLGGAGLGGAG 758
QY 414 QGCTGGLGGGCGVAGGLGGGCGAGAAAAAGAGCGGCGYGV-----GSGASA 457
Db 759 GLGAGGLGAGGLGAGG-GVSPAAAAAATAAAYAGLGGVPGFAGAGVPGFAGAG 808

RESULT 14
GRP_ARATH STANDARD; PRT; 338 AA.
ID GRP_ARATH
AC P27483;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE Glycine-rich cell wall structural protein precursor.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=92003708; PubMed=1912511;
RA Quigley F., Villiot M.L., Mache R.;
RT "Nucleotide sequence and expression of a novel glycine-rich protein
gene from Arabidopsis thaliana."
RL Plant Mol. Biol. 17:949-952(1991).
CC -|- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
```

```
CC -----
CC -|- SUBCELLULAR LOCATION: Cell wall (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X58338; CAA41249.1; -.
DR PIR; S17732; KNUU.
DR Cell wall; Structural protein; Repeat; Signal.
KW SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 338 GLYCINE-RICH CELL WALL STRUCTURAL
FT DOMAIN 21 338 GLY-RICH.
FT SEQUENCE 338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;

Query Match 26.5%; Score 714.5; DB 1; Length 338;
Best Local Similarity 49.0%; Pred. No. 6.6e-27;
Matches 192; Conservative 6; Mismatches 117; Indels 77; Gaps 23;

QY 39 GGAGGGYGGGLGSGAGRGGCGAGAAAAAGAGGGYGGGLGSGAGRGGLGGGAGAAA 98
Db 21 GGGGGGLGGGFGGKFGG-GAGGGFGGAGGGAGGGGAGGGGGGAGGAGGAGGAGG 75
QY 99 AAGGVGGGLGGGCGAGGAGAAAAAGAGGAGGGYGGGLGSGAGRGGGGGGAGAAAAAG 158
Db 76 -GGAG-GAGLG-GAGGGLG-----GGHG-GGIGG-----GAGGAGGGLGGHGGG 120
QY 159 GAGGGYGGGLGSGAGRGGLGGGCGAGAAAAAGAGGAGGGYGGGLGSGAGGGGAGG 218
Db 121 GAGGGGGGLG-----GGIGG-----GAGGAGGGGGLGGHGGGAGGAGGAGG 169
QY 219 GAGRGGLGCGAGAAAAAGAGAGCGGLGG-GAGAAAAAGAGCGGGYGGGLGSGAGRG 277
Db 170 -----GGHGGGICGAGGGGGLGGICGGAGGAGGGGAGGGGAGGAGGAGG 216
QY 278 GAGAAAAAGAGGGYGGGCGAGGGYGGGLGSGAGRGGLGGGCGAGAAAAAGAGGAG 337
Db 217 GGGFGGAGGGLGGGAGGGTG---GGFGG-----GAGGCGAGGAGGG-----FGGAGGG 263
QY 338 GLGCGGAGAAAAAGAGGAGGGGLGGGCGAGCGAGAAAAAGAGGGYGGGLGSGAGR 397
Db 264 AGGFGGAGGAGGAGGAG-GGFGG-GAGGGHG-----GGV-GGGFGG-----GSG- 305
QY 398 GGCGAGAAAAAGAGGAGGGTGLGGGCGVAGG 429
Db 306 GGFGGGAGGAGGAG-GGFGGGGGAGGGGFGG 336

RESULT 15
LORI_MOUSE STANDARD; PRT; 481 AA.
ID LORI_MOUSE
AC P18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE Loricrin.
DE Loricrin.
GN LOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90275605; PubMed=2190691;
RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA Cheng C., Licht U., Bisher M.E., Steven A.C., Steinart P.M.,
RA Yupa S.H., Roop D.R.;
RT "Identification of a major keratinocyte cell envelope protein,
RT loricrin.";
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 30.992 Seconds
(without alignments)
3509.535 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700
Sequence: 1 MASMTGQQMGRIIRIRGGY.....GLSGCDVLIQALLGHHHHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2555	94.6	544	5	O46171
2	2313	85.7	617	5	O46172
3	2290	84.8	644	5	O8WSW4
4	1570.5	58.2	648	5	O9BIU7
5	1405.5	52.1	691	5	O9BIU3
6	1397	51.7	988	5	O17434
7	1305.5	48.4	854	5	O9BIU4
8	1256	46.5	447	5	O9BIV1
9	1099	40.7	636	5	O16987
10	1042.5	38.6	563	5	O9BIT5
11	1025.3	38.0	1933	5	O9BIT7
12	1003	37.1	360	5	O9BIU0
13	968.5	35.9	2639	5	O76786
14	942.5	34.9	1715	16	O8VIZ0
15	941	34.9	2655	5	O964F4
16	935	34.6	1884	5	O9NHW2

17	927	34.3	1079	16	O53557	O53557 mycobacteri
18	926	34.3	1217	16	O8VIX9	O8VIX9 mycobacteri
19	921	34.1	871	5	O44358	O44358 nephila cla
20	912.5	33.8	1489	16	O53559	O53559 mycobacteri
21	900.5	33.4	253	5	O9BIT4	O9BIT4 nephila sen
22	900	33.3	2249	5	O9NHW4	O9NHW4 nephila cla
23	893	33.1	1384	16	O8VIZ1	O8VIZ1 mycobacteri
24	892	33.0	1381	16	O53552	O53552 mycobacteri
25	890.5	33.0	738	5	O02402	O02402 pinctada fu
26	888.5	32.9	444	5	O9BIU6	O9BIU6 argiope tri
27	881	32.6	912	5	O9BIT2	O9BIT2 plectreureys
28	880	32.6	767	16	O53435	O53435 mycobacteri
29	877.5	32.5	1468	5	O9GUB5	O9GUB5 galleria me
30	873	32.3	1002	5	O9BIU8	O9BIU8 argiope tri
31	868.5	32.2	1665	16	O53215	O53215 mycobacteri
32	866.5	32.1	399	5	O9BIT8	O9BIT8 latrodectus
33	862	31.9	907	5	O44359	O44359 nephila cla
34	858	31.8	853	16	O53439	O53439 mycobacteri
35	858	31.8	1408	16	O8VK17	O8VK17 mycobacteri
36	857.5	31.8	882	16	O53845	O53845 mycobacteri
37	857	31.7	879	16	O8VKD2	O8VKD2 mycobacteri
38	852	31.6	651	5	O9BIU9	O9BIU9 argiope tri
39	850.5	31.5	233	5	O9BIT6	O9BIT6 nephila mad
40	850	31.5	749	16	O53844	O53844 mycobacteri
41	847	31.4	626	5	O9NHW1	O9NHW1 nephila mad
42	847	31.4	1329	16	O06810	O06810 mycobacteri
43	843.5	31.2	714	16	O53556	O53556 mycobacteri
44	836	31.0	1538	16	O53395	O53395 mycobacteri
45	835.5	30.9	1507	16	O8VJ23	O8VJ23 mycobacteri

ALIGNMENTS

RESULT 1

O46171	ID	O46171	PRELIMINARY;	PRT;	544 AA.
AC	O46171	01-JUN-1998	(TrEMBLrel. 06, Created)		
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)			
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)			
DE	Spidroin 1 (Fragment).				
OS	Nephila clavipes (orb spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
OC	Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.				
OX	NCBI_TaxID=6915;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90384959; PubMed=2402494;				
RA	Xu M., Lewis R.V.;				
RT	"Structure of a protein superfiber: spider dragline silk."				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).				
RN	[2]				
RP	SEQUENCE OF 449-544 FROM N.A.				
RX	MEDLINE=98148687; PubMed=9487707;				
RA	Arcidiacono S., Mello C., Kaplan D., Cheley S., Bayley H.;				
RT	"Purification and characterization of recombinant spider silk				
RT	expressed in Escherichia coli."				
RL	Appl. Microbiol. Biotechnol. 49:31-38(1998).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98318805;				
RA	Beckwith R., Arcidiacono S., Stote R.;				
RT	"Evolution of repetitive proteins: spider silks from Nephila clavipes				
RT	(Tetragnathidae) and Araneus bicentenarius (Araneidae)."				
RL	Insect Biochem. Mol. Biol. 28:121-130(1998).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RA	Beckwith R.;				
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; U20329; AAC38957.1; -.				
DR	HSSP; P10969; IWGT.				
FT	NON_TER				

OX	NCBI_TaxID=156846;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=21179804; PubMed=11283372;
RX	Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT	"Extreme Diversity, Conservation, and Convergence of Spider Silk
RT	Fibroin Sequences";
RL	Science 291:2603-2605(2001).
DR	EMBL; AF305269; AAK30598.1; .
DR	InterPro; IPR000217; Tubulin.
DR	PROSITE; PS00227; TUBULIN; UNKNOWN_1.
FT	NON_TER 1
FT	NON_TER 854
SQ	SEQUENCE 854 AA; 71881 MW; 5C52F1B0AE31C6C2 CRC64;
	Query Match 48.4%; Score 1305.5; DB 5; Length 854;
	Best Local Similarity 48.6%; Pred. No. 3 6e-59;
	Matches 382; Conservative 29; Mismatches 98; Indels 277; Gaps
Qy	6 GGQGMGRIRYGGGLGOGAGOGAGAAAAAAGAG-----QGYYGGLGSQAG--555
Dd	: :
Dd	64 GGAGSGGGYGCGGGLGYGOGAGACAAAATAAGAGSGSGDYGDDGGLGYG-QGSGAG 122
Qy	56 -----RGGGA-----CAAAAAGGA--GGGY-----GGLG-80
Dd	: : : : : :
Dd	123 SATAPAAAGSGFGCGGFCNGRGKAYQSAGAGVGAATAAAGAGSGGGYGDDGGLG 182
Qy	81 -SOGAG-----RGLG--GOGAG-----AAAAAGVG-----104
Dd	: : : : : :
Dd	183 YGOGAGAAASAAGGDGYEQGGYGNQGLGSFGOGAGAGAAAASAGAGSGRRGYD 242
Qy	105 QGGLGGOGAGQGAGAAAAAAGAGOGGYGGLGSQAGRGSG--GQAGAA-AAAAGGA--160
Dd	: : : : : :
Dd	243 QGGLGYGOGAGAGAAASAAGGGDGYGQGYGPDG--GRGGYGGSGAGSATAAAGGAGCF 301
Qy	161 QGGY--GGLGSOGAGRGLG-CQAGAAAAAAGAG-----QGTYGGL-----202
Dd	: : : : : :
Dd	302 GGGYGQGGYGNQG-GLGSFGOGAGAGAAAASAGAGSGRRGYGDDGGLGYGQAGAG 360
Qy	203 -----GGOGAGGGY--GGLGS--QGAG-----RG 224
Dd	: : : : : :
Dd	361 AAAAAGGDGYGQGGYGNQGLGSFGOGAGAGAAAASAGAGSGRRGYGQGGYGNQGG 420
Qy	225 LG--GOGAG-----AAAAGAG-----QGGLG--GQAGAAAAAAGGA---GQGG 263
Dd	: : : : : :
Dd	421 LGSFGOGAGAGAAAASAGAGSGRRGYGDDGSLGYGOGAGSGAAAAAGGGDGYGQGG 480
Qy	264 YGGLGSOGAGRGQAGAAAAAAGGAGOGGYGQOGAGQGGY---GGLGS--QGAG-----313
Dd	: : : : : :
Dd	481 YGNQG--GLGSFGOGAGAGAAAASAGAGSGRRGYGQGGYGNQGLGSFGOGAGAGAA 538
Qy	314 -----RGLG--GOCAGAAAAAAGGA---GOGGLG-----GOGA 344
Dd	: : : : : :
Dd	539 AASAGAGSGRRGYGDDGGLGYGQAGAGAAAASAGGGDGYGQGGYGNQGVGSGQA 598
Qy	345 AAAAA---AAGGAG-----QGGLGGOGAGOGAGAAAAAAG--GVRGGYGGGLGS 392
Dd	: : : : : :
Dd	599 GAGAATSAAGGAGSGRRGYGEOGGLGY--GOGAGAASTAAGCGDGYGQGGYGNQGG 656
Qy	393 QGA-GRG-GOGAGAA-AAAAGGA-----GGTGTGGL-----GGQVGA 427
Dd	: : : : : : :
Dd	657 RGSYGQGSAGAGAAVAAGGAVSQGGYDGEGGQGGYGGSGAGAAVAASGGTGAGQ 716
Qy	428 GGLG-----GOGAG--AAAA-----VCAGOGGYG----CVGSCASANA--461
Dd	: : : : : :
Dd	717 GGYGSGSQAGYGOAGAFRAAAATAAAGAGGGGGYGGGGYGGGTGAGGASSAGLS 776
Qy	462 -----ASRLSPQASSRVSSANLVSAGPTNSAALSTISNVVSIQIGSNPLGSCDV 515
Dd	: : : : : :
Dd	777 VTGVNMVSRLLSPPEASRVSSVSSLVSNQVNDALPSIINLSSISASATTASDCVE 836
Qy	516 LIQALL 521
	: :

Db 1762 ASGAGSG 1769

Search completed: December 18, 2002, 16:20:59
Job time : 36.9992 secs

This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 22,7682 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEIYNKDGKVDLYKAVGL.....NKLGVGSDTVAIVYQFA 341

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1809	100.0	341	22	AA82612
2	1805	99.8	362	20	AA42550
3	1792	99.1	340	17	AAW92998
4	1051	58.1	367	20	AA42549
5	1051	58.1	367	20	AA34058
6	1051	58.1	367	21	AA57356
7	1035	57.2	377	22	AA69890
8	1033	57.1	367	23	ABB08211
9	1029	56.9	377	20	AA34057
10	1029	56.9	377	21	AA57355

11	972	53.7	573	22	ABG28703	Novel human diagno
12	957.5	52.9	328	22	ABG25541	Novel human diagno
13	957.5	52.9	759	22	ABG25868	Novel human diagno
14	957.5	52.9	759	22	ABG29793	Novel human diagno
15	946	52.3	366	23	AB806345	C. bifermentans DP
16	942	52.1	323	23	AAE17874	Sequence of ompK36
17	858.5	47.5	323	20	AAV34056	E. coli outer memb
18	858.5	47.5	323	21	AA57354	E. coli outer memb
19	752	41.6	393	22	ABG25542	Novel human diagno
20	655	36.2	145	22	AAU20470	Human secreted pro
21	623.5	34.5	440	22	ABG15272	Novel human diagno
22	482	26.6	501	22	ABG25865	Novel human diagno
23	377.5	20.9	518	22	ABG18157	Novel human diagno
24	362.5	20.0	122	22	AA25524	Human protein sequ
25	331.5	18.3	1262	22	ABG29003	Novel human diagno
26	276	15.3	89	18	AAW27781	Amino acid sequenc
27	267.5	14.8	702	22	ABG18349	Novel human diagno
28	267.5	14.8	716	22	ABG25880	Novel human diagno
29	267.5	14.8	924	22	ABG25166	Novel human diagno
30	247.5	13.7	353	22	AAE09801	Aeromonas hydrophi
31	247.5	13.7	373	22	AAE09799	Aeromonas hydrophi
32	170.5	9.4	353	20	AA26025	OmpH protein of H.
33	162.5	9.0	353	20	AA26023	OmpH protein of Pa
34	162	9.0	353	11	AAE06038	Class II outer mem
35	157	8.7	342	16	AAE07063	Mature class 2 por
36	157	8.7	342	18	AAW21742	Neisseria meningit
37	157	8.7	363	16	AAW70764	Fusion Class 2 por
38	157	8.7	363	18	AAW21743	Neisseria meningit
39	152.5	8.4	343	20	AA26024	OmpH protein of Pa
40	150.5	8.3	313	18	AAW21744	Neisseria meningit
41	147.5	8.2	309	16	AAW70762	Meningococcal grou
42	147.5	8.2	309	18	AAW21741	Neisseria meningit
43	144.5	8.0	341	11	AAE07043	P3 gene product of
44	140	7.7	372	16	AAE69607	Conococcal porin-5
45	138.5	7.7	342	16	AAE66879	H. Influenzae prot

ALIGNMENTS

RESULT 1
AA82612

ID AA82612 standard; Protein; 341 AA.

XX AA82612;

XX 02-OCT-2001 (first entry)

DE E. coli outer membrane protein ompF.

XX Outer membrane protein; ompF; structural protein; purification.

XX Escherichia coli.

XX WO200153333-A1.

XX 26-JUL-2001.

XX 01-NOV-2000; 2000WO-US30086.

XX 20-JAN-2000; 2000US-0490291.

PA (MELL/) MELLO C M.

PA (ARCI/) ARCIDIACONO S.

PA (BUTL/) BUTLER M M.

XX (USSA) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

XX WPI; 2001-483136/52.

XX Recovering structural polypeptides in a biological sample, useful for
PT purifying and spinning spider silks and other structural proteins,

PT comprises treating the sample containing the polypeptides with an acid

PS Claim 2; Page 43-44; 49pp; English.

CC The present sequence is that of the Escherichia coli ompF outer
CC structural proteins. The invention provides methods for purifying
CC cells or other biological samples (such as non-recombinant
CC derived cells), and enrich the purity and yields of structural
CC proteins by hydrolysing many of the macromolecules while leaving
CC the structural proteins intact. In the present case, lyophilised
CC pellets of E. coli were lysed in 9.2 N valeric acid for 1 hr at
CC room temperature. The cell lysate was clarified by centrifugation
CC and applied to an SDS polyacrylamide gel for electrophoresis. The
CC ompF protein was then blotted onto a nitrocellulose membrane for
CC N-terminal sequencing. The simple extraction procedure yielded
CC ompF in approximately 8% purity. The new method has the
CC following advantages over prior art: it involves fewer steps,
CC requires less time and smaller volumes of reagents, results in
CC better recovery of protein at higher purity (70-99%), is easy to
CC scale up, and fibres can be spun in an environmentally benign
CC solution reducing hazardous waste accumulation and cost.

XX Sequence 341 AA;

Query Match 100.0%; Score 1809; DB 22; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.1e-147;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ABYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
DB 1 ABYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYSVYDALGYDMLPEFGG 120
DB 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYSVYDALGYDMLPEFGG 120
QY 121 DPAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSGDVGSGSISY 180
DB 121 DPAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSGDVGSGSISY 180
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240
DB 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQFA 341
DB 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQFA 341

RESULT 2
ID AAY42550
XX AAY42550 standard; Protein; 362 AA.

AC AAY42550;

XX 20-DEC-1999 (first entry)

DE E. coli wild-type ompF protein.

KW Bacteria; attenuation; deletion; mutant; vaccine; immune response;
KW Gram negative; infection; diarrhoea; food poisoning; typhoid;
KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.

OS Escherichia coli.

PN W09949026-A1.

XX

PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-GB00935.
XX
PR 25-MAR-1998; 98GB-0006449.
XX
PA (PEPT-) PEPTIDE THERAPEUTICS LTD.
XX
PI Chatfield SN;
XX
DR WPI; 1999-580447/49.
DR N-PSDB; AAZ22887.
PT New attenuated bacteria useful as vaccines for protecting against
PT infections
PS Disclosure; Page 63-64; 69pp; English.
XX

CC This sequence represents the E. coli wild-type ompF protein. The coding
CC sequence of the ompF gene was removed via PCR using primers TT1-TT4
CC (AAZ22889-22893) to produce a non-reverting deletion mutation. The
CC mutant ompF gene (AAZ22888) was used in the production of a bacterium
CC attenuated by a non-reverting mutation in each of the ompF gene, the aroC
CC gene (AAZ22883, AAZ22884), and the ompC gene (AAZ22885, AAZ22886). The
CC mutant bacteria provide immunogenic activity with reduced virulence and
CC thus can be used as a vaccine for raising an immune response against a
CC variety of bacteria in a mammalian host. Such vaccines can provide
CC protection against e.g., E. coli (a cause of diarrhoea in humans),
CC Salmonella typhimurium (the cause of salmonellosis in several animal
CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause
CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis
CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in
CC cattle, especially of new-born calves), Haemophilus influenzae (a cause
CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoeae), Yersinia
CC enterocolitica (the cause of a spectrum of disease in humans ranging
CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis
CC (the cause of whooping cough), and Brucella abortus (a cause of abortion
CC and infertility in cattle and a condition known as undulant fever in
CC humans).

XX Sequence 362 AA;

Query Match 99.8%; Score 1805; DB 20; Length 362;
Best Local Similarity 100.0%; Pred. No. 2.6e-147;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEIYNKDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
DB 23 AEIYNKDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 82
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYSVYDALGYDMLPEFGG 120
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYSVYDALGYDMLPEFGG 142
QY 121 DPAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSGDVGSGSISY 180
DB 143 DPAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSGDVGSGSISY 202
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240
DB 203 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 262
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 263 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 322
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362

RESULT 3
AAW92998

```
ID AAW92998 standard; protein; 340 AA.
XX
AC AAW92998;
XX
DT 19-MAY-1999 (first entry)
DE
DE E. coli OmpF porin protein.
XX
KW OmpF; porin; amphiphilic alpha-structure; beta-structure; GLUT;
KW membrane protein; glucose transporter protein; function.
XX
OS Escherichia coli.
XX
PN WO9618957-A1.
XX
PD 20-JUN-1996.
XX
PF 13-DEC-1995; 95WO-US16126.
XX
PR 14-DEC-1994; 94US-0355844.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Cheung M, Czegledy F, Fischbarg J, Iserovich P;
PI Li J;
PI WPI; 1996-300839/30.
XX
XX Predicting tendency to form amphiphilic alpha and beta structures -
PT using a novel algorithm to calculate values for subsequent graphical
PT analysis to predict protein structure
XX
PS Disclosure; Fig 2; 106pp; English.
XX
CC This invention describes a novel method for predicting the tendency of
CC a protein to form either: (a) an amphiphilic alpha-structure or (b) an
CC amphiphilic beta-structure. The methods are used particularly for
CC predicting the structure of membrane proteins such as glucose transporter
CC proteins (GLUTs). They can be used to discern the function of proteins.
CC They can also be used for the rational design or identification of
CC compounds which interact with the proteins or to engineer proteins having
CC particular structures. This sequence represents an Escherichia coli
CC OmpF porin which is used to illustrate the method of the invention.
XX
SQ Sequence 340 AA;
Query Match 99.1%; Score 1792; DB 17; Length 340;
Best Local Similarity 99.4%; Pred. No. 3.2e-146;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGYGGNDMTYARLFGKGETQINSDLTGYGQ 60
DB 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGYGGNDMTYARLFGKGETQINSDLTGYGQ 60
QY 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLVADYGSFDRYGRNYGVVYDALGYTDLPEFGG 120
DB 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLVADYGSFDRYGRNYGVVYDALGYTDLPEFGG 120
QY 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDTARRSNGDVGGSISY 180
DB 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDTARRSNGDVGGSISY 180
QY 181 EYEGFGIVGAYGAADRNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
DB 181 EYEGFGIVGAYGAADRNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
QY 241 TNKFTNTSGFANKTQDVLVLAQYQFFGLRPSIATYTKSKAKDVEGIGDVLVNFVEGAT 300
DB 241 TNKFTNTSGFANKTQDVLVLAQYQFFGLRPSIATYTKSKAKDVEGIGDVLVNFVEGAT 300
QY 301 YFYNKNMSTVYDIINQIDSDNKLGVGSDDTVAAGIVYQF 340
DB 301 YFYNKNMSTVYDIINQIDSDNKLGVGSDDTVAAGIVYQF 340
```

```
RESULT 4
AAI42549
ID AAY42549 standard; Protein; 367 AA.
XX
AC AAY42549;
XX
DT 20-DEC-1999 (first entry)
DE
DE E. coli wild-type ompC protein.
XX
XX Bacteria; attenuation; deletion; mutant; vaccine; immune response;
KW Gram negative; infection; diarrhoea; food poisoning; typhoid;
KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.
XX
OS Escherichia coli.
XX
PN WO9949026-A1.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-GB00935.
XX
PR 25-MAR-1998; 98GB-0006449.
XX
PA (PEPT-) PEPTIDE THERAPEUTICS LTD.
XX
XX Chatfield SN;
XX
XX WPI; 1999-580447/49.
XX
XX N-PSDB; AA222885.
XX
XX New attenuated bacteria useful as vaccines for protecting against
XX infections -
XX Disclosure; Page 57-58; 69pp; English.
XX
CC This sequence represents the E. coli wild-type ompC protein. The coding
CC sequence of the ompC gene was removed via PCR using primers TT7-TT10
CC (AA222893-422896) to produce a non-reverting deletion mutation. The
CC mutant ompC gene (AA222886) was used in the production of a bacterium
CC attenuated by a non-reverting mutation in each of the ompC gene, the aroC
CC gene (AA222883, AA222884), and the ompF gene (AA222887, AA222888). The
CC mutant bacteria provide immunogenic activity with reduced virulence and
CC thus can be used as a vaccine for raising an immune response against a
CC variety of bacteria in a mammalian host. Such vaccines can provide a
CC protection against e.g., E. coli (a cause of diarrhoea in humans),
CC Salmonella typhimurium (the cause of salmonellosis in several animal
CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause
CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis
CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in
CC cattle, especially of new-born calves), Haemophilus influenzae (a cause
CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia
CC enterocolitica (the cause of a spectrum of disease in humans ranging
CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis
CC (the cause of whooping cough), and Brucella abortus (a cause of abortion
CC and infertility in cattle and a condition known as undulant fever in
CC humans).
XX
SQ Sequence 367 AA;
Query Match 58.1%; Score 1051; DB 20; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.5e-82;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
QY 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGYGGNDMTYARLFGKGETQINSDLTGYGQ 60
DB 22 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGYGGNDMTYARLFGKGETQINSDLTGYGQ 76
QY 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLVADYGSFDRYGRNYGVVYDALGYTDLPEFGG 120
DB 77 WEYQIQGNNAEN---ENNSWTRVAFAGLVADYGSFDRYGRNYGVVYDTSWTDVLPFEGG 133
```



```

XX SQ Sequence 367 AA;
Query Match 58.1%; Score 1051; DB 21; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.5e-82;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGETQINSLDTGYGQ 60
DB 22 AEVYNKGNKLDLYGKVDGLHYFSDNSAK-----DGDSYARLGFKGETQINDQLTGYGQ 76
QY 61 WEYNFQGNNSGADAQTCNKTRLAFAGLKADYGVSGFYGRNYGVVYDALGYTMDLPEFGG 120
DB 77 WEYQIQGNSAEN---ENNSTWTRVAFAGLKFDQYGVSGFYGRNYGVVYDVTWTDVLPFEGG 133
QY 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYGLKNERDT-----A 166
DB 134 DTYSGNDFMQQRGNFGATYRTDFFGLVDGLNFAVQYQGNKNGSPSGEGFTSGVTNNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFGIVGAYGAADRTNLOE-AOPLGNGKKAQOWATGLKAYDANNI 225
DB 194 LRQNGDVGGSITDYEGFGIGGAISSSKRTDAQNTAAAYICNGDRAETTYGGLYDANNI 253
QY 226 YLAANYGETRNATPTNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDY-E 284
DB 254 YLAAQYQTYNATRV-----GSLGWANKAQNEFAVQYQDFGLRPSIAYLQSKGNLGR 308
QY 285 GIGDVLNVTPEVCATYFKNKMSTYVDYIINQIDSNKL-----GVGSDDTVAVGIVYQF 340
DB 309 GYDEDDILKYVDGATYFKNKMSTYVDYKINLLD-DNQETRDAGINTDNIVALGLVYQF 367

RESULT 7
AAG98980
ID AAG98980 standard; Protein; 377 AA.
XX
AC AAG98980;
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related protein sequence SEQ ID NO:450.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
OS Escherichia coli.
XX
PN WO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zyskind J;
XX
DR WPI; 2001-335933/35.
DR N-PSDB; AAH84651.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors
XX
PS Claim 19; Page 505-506; 522pp; English.
XX
CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
CC growth and proliferation related proteins given in AAG99078 and AAG98830
CC to AAG98999. (I) can be used as potential targets for the generation of
CC new antimicrobial agents, and for identification of compounds which

```

```

CC interact with the gene products of (I). In addition the expression of
CC (I) and the purification of the proteins, the purified proteins can be
CC used to generate reagents and screen small molecule libraries or other
CC candidate compound libraries for compounds that can be further developed
CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (I) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable
CC method by which to identify the causative agents of a bacterial
CC infection. Also, antibodies generated against proteins translated from
CC mRNA transcribed from proliferation-required sequences can also be used
CC to screen for specific microorganisms that produce such proteins in a
CC species-specific manner. AAH84371 and AAH84670 represent sequencing
CC primers used in the isolation of E. coli growth and proliferation
CC related sequence, which are used in an example from the present
CC invention.
XX
SQ Sequence 377 AA;
Query Match 57.2%; Score 1035; DB 22; Length 377;
Best Local Similarity 56.9%; Pred. No. 6.3e-81;
Matches 209; Conservative 42; Mismatches 78; Indels 38; Gaps 9;

QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGETQINSLDTGYGQ 60
DB 22 AEVYNKGNKLDLYGKVDGLHYFSDNSAK-----DGDSYARLGFKGETQINDQLTGYGQ 76
QY 61 WEYNFQGNNSGADAQTCNKTRLAFAGLKADYGVSGFYGRNYGVVYDALGYTMDLPEFGG 120
DB 77 WEYNIQANTESSKNQSW--TRLAFAGLKAFADYGVSGFYGRNYGVYDTEGWTMLPEFGG 134
QY 121 DT-AYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYGLKNE-----RDTAR 167
DB 135 DSYTNADNFMGTGRANGVATYRNTDFFGLVNLNFAVQYQGNNGASNGEQETNNGRD-VR 193
QY 168 RNSGDVGGSISYEY-EGFGIVGAYGAADRTNLOEAOPLGNGKKAQOWATGLKAYDANNIY 226
DB 194 HENGDGWGLSTTYDLGWFSGAGAAATSSDRTNDQVNHNTAAGDKADAWTAGLKAYDANNIY 253
QY 227 LAANYGETRNATPTNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVE-- 284
DB 254 LATWSETRNTPTFGD---SDYAVANKTQNFVTAQYQDFGLRPAVSFLSKGRDLHAA 310
QY 285 -----GIGDVLNVTPEVCATYFKNKMSTYVDYIINQIDSNKL-----GVGSDDTVA 333
DB 311 GGADNPAGVDDKDLVKYADIGATYFKNKMSTYVDYKINLLDEDDSFYAANGISTDDIVA 370
QY 334 VGIVYQF 340
DB 371 LGLVYQF 377

RESULT 8
ABB08211
ID ABB08211 standard; protein; 367 AA.
XX
AC ABB08211;
XX
DT 08-APR-2002 (first entry)
XX
DE Escherichia coli outer membrane protein C (OmpC) precursor.
XX
KW OmpC; outer membrane protein C; Crohn's disease; immunoglobulin A; IgA;
KW antiinflammatory; immune response.
XX
OS Escherichia coli.
XX
PN WO200189361-A2.
XX
PD 29-NOV-2001.
XX
PF 17-MAY-2001; 2001WO-US16032.
XX

```

```
PR 19-MAY-2000; 2000US-0575061.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PA (REGC ) UNIV CALIFORNIA.
XX
PI Targan SR, Braun J, Sutton CL;
XX WPI; 2002-075333/10.
XX
PT Diagnosing Crohn's disease in a subject comprises detecting the
PT presence of immunoglobulin A anti-outer membrane protein C antibodies -
XX
PS Claim 3; Fig 5; 55pp; English.
XX
CC The sequence represents the E. coli outer membrane protein C (OmpC)
CC precursor. The invention relates to a novel method for diagnosing Crohn's
CC disease in a subject. The method comprises detecting immunoglobulin (Ig)A
CC OmpC antibodies in the subject, where the presence of the IgA anti-OmpC
CC antibodies indicates that the subject has Crohn's disease. The proteins
CC of the invention have antiinflammatory activity. The method of the
CC invention is an immunological response elicitor. The method is useful for
CC diagnosing Crohn's disease in a subject and administering an OmpC antigen
CC is useful for inducing tolerance in a patient with Crohn's disease.
XX
SQ Sequence 367 AA;
Query Match 57.1%; Score 1033; DB 23; Length 367;
Best Local Similarity 58.3%; Pred. No. 9e-81;
Matches 210; Conservative 45; Mismatches 71; Indels 34; Gaps 8;
QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMYARLGFKGETQINSDLTG YGO 60
DB 22 AEIYNKDGKNDLYGKVDGLHYFS-----DNKDVGDGDTYMLRGLFKGETQVTDQLTG YGG 76
QY 61 WEYNFQGNSEGADAQTGNKTRLAFAGLVADVGSDFDYGRTYGVVYDALGYTDMLEPFGG 120
DB 77 WEYIQGNSAEN---ENNNTVRFAFLAGLQDVGSDYGRNTYGVVYDVTSTWTDVLEPFGG 133
QY 121 DTAYSDFFVGRGVGATYRNNSFFGLVDGLNFAVQYLGKNERDT-----A 166
DB 134 DTYGSDNFNQQRGNFGATYRNTDFFGLVDGLNFAVQYQKGNPSGEGFTSGVTNNGRDA 193
QY 167 RNSGDVGSGSISYEYEGFIVGAYCAADRTNLQE-AQPLGNGKKAQWATGLKYDANNI 225
DB 194 LRQNGDVGSGSITYDEYEGFIVGAYCAADRTNLQE-AQPLGNGKKAQWATGLKYDANNI 253
QY 226 YLAANYGETRNATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRPSIATYKSKADV-E 284
DB 254 YLAANYQTYNATRV-----GSLGWANKAQNFENVAQYQDFGLRPSLAYLQSKGNLGR 308
QY 285 GIGDVLVNYFEVGATYTFYFNKNMSTYVDYIINQIDSDNKL-----GVGSDTVAVGIVYQF 340
DB 309 GYDDEDILKYVDVGTATYTFYFNKNMSTYVDYIINQIDSDNKL-DNQFTRDAGINTDNIVALGLVYQF 367
RESULT 9
AAV34057
ID AAV34057 standard; protein; 377 AA.
XX
AC AAV34057;
XX
XX 23-NOV-1999 (first entry)
XX
DE E. coli outer membrane protein F precursor.
XX
KW Ulcerative colitis; histone; H1-like antigen; porin antigen;
KW Bacteroides antigen; IBD; PANCA; inflammatory bowel disease;
KW diagnosis; perinuclear anti-neutrophil cytoplasmic antibody;
KW outer membrane protein F precursor.
XX
OS Escherichia coli.
XX
PN W09945955-A1.
```

```
XX 16-SEP-1999.
PD
XX
XX 12-MAR-1999; 99WO-US03492.
PF
XX
XX 12-MAR-1998; 98US-0041889.
PR
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Braun J, Cohavy O;
XX WPI; 1999-551215/46.
DR
XX
XX Use of histone H1, porin or Bacteroides antigens as targets for the
XX diagnosis, prevention and treatment of ulcerative colitis -
PT
PT Claim 8; Fig 11; 134pp; English.
XX
XX The invention provides a method for the diagnosis, prevention and
XX treatment of ulcerative colitis (UC) using histone H1-like antigen, a
XX porin antigen or a Bacteroides antigen as a target antigen. The novel
XX method of diagnosing UC in a subject suspected of having inflammatory
XX bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
XX (2) contacting the sample with a histone H1-like antigen, or perinuclear
XX anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
XX complex of the histone H1-like antigen, or the PANCA-reactive fragment,
XX and antibody to the histone H1-like antigen; and (3) detecting the
XX presence or absence of the complex; where the presence of the complex
XX indicates that the subject has UC. The PANCA-reactive histone H1-like
XX antigen, porin antigen and Bacteroides antigen are useful in the
XX diagnosis, prevention and treatment of UC. The methods can also be used
XX for identifying agents useful for treating UC. The present sequence
XX represents a E. coli outer membrane protein F precursor.
SQ Sequence 377 AA;
Query Match 56.9%; Score 1029; DB 20; Length 377;
Best Local Similarity 56.9%; Pred. No. 2.1e-80;
Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;
QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMYARLGFKGETQINSDLTG YGO 60
DB 22 AEIYNKDGKNDLYGKVDGLHYFS-----DGDQSYARLGFKGETQINDQLTG YGO 76
QY 61 WEYNFQGNSEGADAQTGNKTRLAFAGLVADVGSDFDYGRTYGVVYDALGYTDMLEPFGG 120
DB 77 WEYINQANTTESKNSW--TRLAFAGLVADVGSDFDYGRTYGVVYDALGYTDMLEPFGG 134
QY 121 DT-AYSDDFVGRGVGATYRNNSFFGLVDGLNFAVQYLGKNE-----RDTAR 167
DB 135 DSYTNADNFMTCRANGVATYRNTDFFGLVNLNFAVQYQGNNEGASNGQEGTNNGRD-VR 193
QY 168 RNSGDVGSGSISYEY-EGFIVGAYCAADRTNLQE-AQPLGNGKKAQWATGLKYDANNI 226
DB 194 HENGDCWGLSTTYDLMGMFSGAAGATSSDRTNDQVNHNTAAGDKADAWTAGLKYDANNI 253
QY 227 LAANYGETRNATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRPSIATYKSKADV-E 284
DB 254 LATMYSETRNMTFFGD---SDYAVANKTONFEVTAQYQDFGLRPAVSLFMSKGRDLHAA 310
QY 285 -----GIGDVLVNYFEVGATYTFYFNKNMSTYVDYIINQIDSDNKL-----GVGSDTVA 333
DB 311 GGADNPAGVDDKDLVKYADIGATYTFYFNKNMSTYVDYIINQIDSDNKL-----GVGSDTVA 370
QY 334 VGVIVYQF 340
DB 371 LGLVYQF 377
RESULT 10
AAV57355
ID AAV57355 standard; protein; 377 AA.
XX
```

AA57355;
 13-JUN-2000 (first entry)
 E. coli outer membrane protein F precursor.
 Ulcerative colitis; inflammatory bowel disease; porin antigen; Mab;
 pANCA; perinuclear anti-neutrophil cytoplasmic antibody;
 histone H1; outer membrane protein F precursor.
 Escherichia coli.
 US6033864-A.
 07-MAR-2000.
 12-MAR-1998; 98US-0041889.
 12-APR-1996; 96US-0057846.
 11-APR-1997; 97US-0837058.
 (REGC) UNIV CALIFORNIA.
 Cohavy O, Braun J;
 WPI; 2000-255695/22.
 Diagnosing ulcerative colitis or susceptibility, by detecting complex
 formation between microbial porin antigen and perinuclear
 anti-neutrophil cytoplasmic autoantibodies -
 Claim 1; Fig 10; 49pp; English.
 The invention provides a method for diagnosing ulcerative colitis in a
 subject suspected of having inflammatory bowel disease. The method
 comprises reacting a patient sample with a porin antigen that is
 immunologically reactive with pANCA (perinuclear anti-neutrophil
 cytoplasmic antibodies) and detecting formation of a Ag-PANCA complex
 as indicative of ulcerative colitis. The method is used to diagnose
 ulcerative colitis or susceptibility to it. The present sequence
 represents a E. coli outer membrane protein F precursor.
 Claim 1; Fig 10; 49pp; English.
 Query Match 56.9%; Score 1029; DB 21; Length 377;
 Best Local Similarity 56.9%; Pred. No. 2.1e-80;
 Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;
 1 AEIYKDKGKVDLYGRAVGLHYFSKGENSYGNGDMYARLFGKGTQINSDLTGYGQ 60
 22 AEIYKDKGKLDLYGKVDGLHYFSKGENSYGNGDMYARLFGKGTQINSDLTGYGQ 76
 61 WEYNFQGNSEGADQGNKTRIAFLAGLKYADYGVSGFDYGRNYGVVYDALGYTDMLPFEGG 120
 77 WEYNIQANTESSKQNSW--TRLAFLAGLKYADYGVSGFDYGRNYGVVYDALGYTDMLPFEGG 134
 121 DT-AYSDFFVGRVGGVATYRNSNFFCLVDGLNFAVQYLGKNE-----RRTAR 167
 135 DSTNADNFMTGRANGVATYRNTDFGLVNLNFAVQYQGNNEGASNGQEGTNGRD-VR 193
 168 RNSGDGVGSGISVEY-EGFIVGAYGAADRNLQEAQPLNGKKAQWATGLKYDANNIY 226
 194 HENGDDGHLSTTYDLGLGFFAGAGAYTSSDNTNDQVNTAAGGDKADWTAGLKYDANNIY 253
 227 LAANYGETRNATPITNKTFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVE-- 284
 254 LATMYSETRNMTPEGD---SDYAVANKTQFVEYTAQYQDFGLRPAVSFLMSKGRDLHAA 310
 285 -----GIGDVLVNYFEVGYATYFNKNMSTYVDYIINOIDSDNKL----GVGSDDTVA 333
 311 GGADNPAGVDKDLVKYADYIGATYTYFNKNMSTYVDYKINLIDEDDSFYAANGISTDDIVA 370
 334 VGIYQF 340

Db 371 LGLVYQF 377
 RESULT 11
 ABG28703
 ID ABG28703 standard; Protein; 573 AA.
 AC ABG28703;
 18-FEB-2002 (first entry)
 Novel human diagnostic protein #28694.
 Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.
 Homo sapiens.
 WO200175067-A2.
 11-OCT-2001.
 30-MAR-2001; 2001WO-US08631.
 31-MAR-2000; 2000US-0540217.
 23-AUG-2000; 2000US-0649167.
 (HYSE-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPI; 2001-639362/73.
 N-PSDB; AAS92890.
 New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity -
 Claim 20; SEQ ID NO 59062; 103pp; English.
 The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. ABG00010-ABG30377 represent novel human
 diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Query Match 53.7%; Score 972; DB 22; Length 573;
 Best Local Similarity 61.0%; Pred. No. 3e-75;
 Matches 188; Conservative 33; Mismatches 71; Indels 16; Gaps 5;
 1 AEIYKDKGKVDLYGRAVGLHYFSKGENSYGNGDMYARLFGKGTQINSDLTGYGQ 60
 28 AEIYKDKGKLDLYGKVKAMHYS-----DNASKDQDSYIRFGKGTQINDLTGYGR 82

Db 715 ATYYFNKNSTYVDYKINLLD-DNQFTRDAGINTDNIVALGLVYQF 759

RESULT 15

ABB06345

ID ABB06345 standard; Protein; 366 AA.

XX ABB06345;

XX 30-MAY-2002 (first entry)

XX C. bifermentans DPH-1 perchloroethylene dehalogenase SEQ ID NO:1.

XX Clostridium bifermentans DPH-1; perchloroethylene dehalogenase;

KW enzyme; decomposing; chlorinated aliphatic compound.

XX Clostridium bifermentans.

XX JP2002017358-A.

XX 22-JAN-2002.

XX 04-JUL-2000; 2000JP-0202729.

XX 04-JUL-2000; 2000JP-0202729.

XX (GIFU-) GIFU DAIGAKUCHO.

XX WPI; 2002-263241/31.

XX N-PSDB; ABL49767.

XX A perchloroethylene dehalogenase derived from Clostridium bifermentans DPH-1, for decomposing chlorinated aliphatic compounds. -

XX Claim 4; Page 11-12; 15pp; Japanese.

XX The present sequence represents perchloroethylene (PCE) dehalogenase derived from Clostridium bifermentans DPH-1. PCE dehalogenase: (a) catalyses the dehalogenising reaction forming trichloroethylene from tetrachloroethylene; (b) is a dimer consisting of two subunits having a molecular weight of 35 kDa; (c) has an optimum temperature of 35 plus degrees Celsius; (d) has an optimum pH of 7.5; (e) has enzymatic activity that is not affected by the addition of metal ions, ethylene diamine tetraacetate, reduced nicotinamide adenine dinucleotide or cyanocobalamin; and (f) has enzymatic activity which is inhibited by the presence of oxygen. The enzyme is used for decomposing chlorinated aliphatic compounds.

XX Sequence 366 AA;

Query Match 52.3%; Score 946; DB 23; Length 366;

Best Local Similarity 55.8%; Pred. No. 2.8e-73;

Matches 191; Conservative 36; Mismatches 85; Indels 30; Gaps 8;

Qy 1 AEIYNKDGKVDLYKAVGLHYFESKNGENSGNGNDMTYARLGFKEGTQINSDLTGYGQ 60

Db 22 AEVYNKDGKVDLYKAVGLHYFESKNGENSGNGNDMTYARLGFKEGTQINSDLTGYGQ 76

Qy 61 WEYNFQGNSEGAQAQTKGNKTRLAFAGLVADVGSDYGRNKGVDYDALGYTDMLPFEGG 120

Db 77 WEYQVQANGTEGDKCDSW--TRLAFAGLVADVGSDYGRNKGVDYDALGYTDMLPFEGG 134

Qy 121 DT-AYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGNKNERDT----- 165

Db 135 DSYTKADNFMTRANGVATYRNTDFGLVDGLNFAVQYLGNKNERDT----- 194

Qy 166 ARRSGDCVCGSISVEYEGFIV--GAYCAADRNLQEAQPLGNGCKKAEQWATGLKYDAN 223

Db 195 MKNNGDGFISSTYDL-GMGVSFGAAYTSSDRINEQVNDSTAGGDKADAWTVGLKYDAN 253

Qy 224 NIYLAANYETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDV 283

Db 254 NIYLATWYSETRNTPYGGSGNSDNTIANKTQNFVTAQYQDFGLRPSVFLSKGKDL 313

Qy 284 EGI----GDVLDLVNTEFGATYFKNKMSYVDYIINQIDSD 321
Db 314 -GVNGSDGDDLVKVASVGATYFKNKMSYVDYIINQIDSD 354

Search completed: December 18, 2002, 16:17:26
Job time : 24.7682 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 : Search time 7.85111 Seconds
(without alignments)
1277.936 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEIYNKDGKVDLYGKAVGL.....NKLGVGSDDTVAIVYQFA 341

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.1	340	2	US-08-355-844-1
2	1792	99.1	340	5	Sequence 1, Appli
3	1051	58.1	367	3	Sequence 1, Appli
4	1029	56.9	377	3	Sequence 30, Appl
5	858.5	47.5	323	3	Sequence 29, Appl
6	157	8.7	342	1	Sequence 28, Appl
7	157	8.7	342	1	Sequence 4, Appli
8	157	8.7	342	1	Sequence 4, Appli
9	157	8.7	342	3	Sequence 4, Appli
10	157	8.7	363	1	Sequence 6, Appli
11	157	8.7	363	1	Sequence 6, Appli
12	157	8.7	363	3	Sequence 6, Appli
13	157	8.7	363	5	Sequence 6, Appli
14	147.5	8.2	309	1	Sequence 2, Appli
15	147.5	8.2	309	1	Sequence 2, Appli
16	147.5	8.2	309	3	Sequence 2, Appli
17	147.5	8.2	309	5	Sequence 2, Appli
18	138.5	7.7	342	4	Sequence 12, Appl
19	138.5	7.7	342	4	Sequence 14, Appl
20	138.5	7.7	342	5	Sequence 12, Appl
21	138.5	7.7	342	5	Sequence 14, Appl
22	138.5	7.7	361	4	Sequence 8, Appli
23	138.5	7.7	361	5	Sequence 8, Appli
24	138.5	7.7	363	4	Sequence 10, Appl
25	138.5	7.7	363	5	Sequence 10, Appl
26	135.5	7.5	350	4	Sequence 14, Appl
27	120.5	6.7	326	4	Sequence 12, Appl

28	114.5	6.3	707	4	US-09-021-560-4	Sequence 4, Appli
29	111	6.1	744	4	US-09-021-560-2	Sequence 2, Appli
30	108.5	6.0	511	1	US-08-480-604A-20	Sequence 20, Appl
31	108.5	6.0	511	2	US-08-405-496A-20	Sequence 20, Appl
32	108.5	6.0	511	4	US-08-915-136-20	Sequence 20, Appl
33	108.5	6.0	511	4	US-08-957-310-20	Sequence 20, Appl
34	108.5	6.0	608	1	US-08-480-604A-21	Sequence 21, Appl
35	108.5	6.0	608	2	US-08-405-496A-21	Sequence 21, Appl
36	108.5	6.0	608	4	US-08-915-136-21	Sequence 21, Appl
37	108.5	6.0	608	4	US-08-957-310-21	Sequence 21, Appl
38	108.5	6.0	609	1	US-08-480-604A-30	Sequence 30, Appl
39	108.5	6.0	609	4	US-08-915-136-30	Sequence 30, Appl
40	108.5	6.0	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	108.5	6.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
42	108.5	6.0	2366	4	US-08-915-136-10	Sequence 10, Appl
43	108.5	6.0	2366	4	US-08-957-310-10	Sequence 10, Appl
44	105	5.8	455	2	US-08-472-172-4	Sequence 4, Appli
45	104	5.7	848	4	US-09-556-877-192	Sequence 192, App

ALIGNMENTS

RESULT 1
US-08-355-844-1
; Sequence 1, Application US/08355844
; Patent No. 5940307
; GENERAL INFORMATION:
; APPLICANT: Fischbarg, Jorge
; APPLICANT: Czegledy, Ferenc
; APPLICANT: Iserovich, Pavel
; APPLICANT: Li, Jun
; APPLICANT: Cheung, Min
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
; TITLE OF INVENTION: STRUCTURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/355,844
; FILING DATE: 14-DEC-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Tang, Henry Y. S.
; REGISTRATION NUMBER: 29,705
; REFERENCE/DOCKET NUMBER: A29927-50/29910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2586
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..340
; OTHER INFORMATION: OmpF porin protein
; US-08-355-844-1

```
Query Match          99.1%; Score 1792; DB 2; Length 340;
Best Local Similarity 99.4%; Pred. No. 3e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGGNDMTYARLGFKEGTQINSDLTGYGQ 60
DB 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGGNDMTYARLGFKEGTQINSDLTGYGQ 60
QY 61 WEYNFQGNSEGADATQGNKTRLAFAGLYADVGSDYGRNRYGVVYDALGYTDMLPFEGG 120
DB 61 WEYNFQGNSEGADATQGNKTRLAFAGLYADVGSDYGRNRYGVVYDALGYTDMLPFEGG 120
QY 121 DTAYSDDDFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNERTARRSNGDVGGSISY 180
DB 121 DTAYSDDDFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNERTARRSNGDVGGSISY 180
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAVTYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAVTYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

RESULT 2
PCT-US95-16126-1
; Sequence 1, Application PC/TU959516126
; GENERAL INFORMATION:
; APPLICANT: Fischbaig, Jorge
; APPLICANT: Czegledy, Ferenc
; APPLICANT: Iserovich, Pavel
; APPLICANT: Li, Jun
; APPLICANT: Cheung, Min
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/355,844
; FILING DATE: 14-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Tang, Henry Y.S.
; REGISTRATION NUMBER: 29,705
; REFERENCE/DOCKET NUMBER: A29927-50/29910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2586
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..340
; OTHER INFORMATION: OmpF porin protein
PCT-US95-16126-1

Query Match          99.1%; Score 1792; DB 5; Length 340;
Best Local Similarity 99.4%; Pred. No. 3e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGGNDMTYARLGFKEGTQINSDLTGYGQ 60
DB 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGGNDMTYARLGFKEGTQINSDLTGYGQ 60
QY 61 WEYNFQGNSEGADATQGNKTRLAFAGLYADVGSDYGRNRYGVVYDALGYTDMLPFEGG 120
DB 61 WEYNFQGNSEGADATQGNKTRLAFAGLYADVGSDYGRNRYGVVYDALGYTDMLPFEGG 120
QY 121 DTAYSDDDFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNERTARRSNGDVGGSISY 180
DB 121 DTAYSDDDFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNERTARRSNGDVGGSISY 180
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAVTYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAVTYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

RESULT 3
US-09-041-889-30
; Sequence 30, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-30

Query Match 58.1%; Score 1051; DB 3; Length 367;
Best Local Similarity 58.9%; Pred. No. 4.1e-90;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYGQ 60
DB 22 AEIYNKGNKLDLYGKVDGLHYFS-----DNKDVGDGQTYMRLFGKGTQVTDQLTGYYGQ 76
QY 61 WEYNFQGNSEGADAOQGNKTRLAFAAGLYADYGSFDYGRNYGVVYDALGYTDLMLPEFGG 120
DB 77 WEYQIGNSAEN---ENNSWTRVAFAGLKFQDVGSFDYGRNYGVVYDVTWTDVLPFEGG 133
QY 121 DTAYSDDFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNERDT-----A 166
DB 134 DTGSDNFMQQRNGFATYRNTDFGLVDGLNFAVQYQKNGNPSGEGFTSGVTNNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFGIYGAAGADRTNLQE-AQPLGNKKAQWATGLKYDANNI 225
DB 194 LRQNGDVGGSITYDEYEGFGIGGAISSKRTDAQNTAAAYIGNGDRAETTYGGLKYDANNI 253
QY 226 YLAANYGETRNPATITNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKARDV-E 284
DB 254 YLAQQTQTYNATRV-----GSLGWANKAQNFAVAQYQDFGLRPSLAYLQSKGNLGR 308
QY 285 GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDTVAVGIVYQF 340
DB 309 GYDDEDILKYVDVGATYFYNKNNSTVYDKINLLD-DNQFTRDAGINTDNIVALGLVYQF 367

RESULT 4
US-09-041-889-29
Sequence 29, Application US/09041889
Patent No. 6033864
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Cohavy, Offer
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-29

Query Match 56.9%; Score 1029; DB 3; Length 377;
Best Local Similarity 56.9%; Pred. No. 4.8e-88;
Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;
QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYGQ 60
DB 22 AEIYNKGNKLDLYGKVDGLHYFS-----DGDSYARLFGKGTQINDOLTGYYQ 76
QY 61 WEYNFQGNSEGADAOQGNKTRLAFAAGLYADYGSFDYGRNYGVVYDALGYTDLMLPEFGG 120
DB 77 WEYNIQANNTESKQNSW--TRLAFAAGLKFADYGSFDYGRNYGVYDIEGWTDLMLPEFGG 134
QY 121 DT-AYSDDFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNE-----RDTAR 167
DB 135 DSTNADNFMTRANGVATYRNTDFGLVNLNFAVQYQNGEASNGOEGTNNGRD-VR 193
QY 168 RNSGDVGGSISYEY-EGFGIVGAYGAADRTNLQEAQPLGNKKAQWATGLKYDANNIY 226
DB 194 HENGDWGLSTYDGLNGFSGAAYTSSDRTNDQVNIHTAAGDKADANTAGLKYDANNIY 253
QY 227 LAANYGETRNPATITNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKAKDVE-- 284
DB 254 LATMYSETRNTTPEGD---SDYAVANKTQNFVTAQYQDFGLRPAVSFLMSKGRDLHAA 310
QY 285 -----GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDDTVA 333
DB 311 GGADNPAGVDDKDLVKYADIGATYFYNKNNSTVYDIYNKINLIDEDDSFYANGISTDDIVA 370
QY 334 VGIVYQF 340
DB 371 LGLVYQF 377

RESULT 5
US-09-041-889-28
Sequence 28, Application US/09041889
Patent No. 6033864
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Cohavy, Offer
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-041-889-28

Query Match 47.5%; Score 858.5; DB 3; Length 323;
Best Local Similarity 57.2%; Pred. No. 3.1e-72;
Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;
QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
Db |||||:|||||:|||||: : :||:|||||:|||||:|||||:|||||:
22 AEVYNKDGKLDLYCKVDGLHYFSNSAK-----DGDQSYARLGFKEGTQINDLTLTGQ 76
QY 61 WEYNFGNNSGADAGTGNKTRLAFLAGLKADYGVDFGRNRYGVVYDALGYTDMLPFEG 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 WEYNIQANNTSSKNQSW--TRLAFLAGLKADYGVDFGRNRYGVVYDIBGWTDMLPFEG 134
QY 121 DR-AVSDDFEFCVRGVGATYRNSNFFGLVGLNFAVQYLGKNE-----RDTRAR 167
Db ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 DSYTNADNFMTRGRANGVATYRNTDFGLVGLNFAVQYGNNEGASNGOEGTNGRDR-VR 193
QY 168 RNSGDSGVGSSISYEX-EGFGIVGAYGAADRTNLQEAQPLNGKKAQEWATGLKYDANNIY 226
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 HENGDSGLSTYDILGMEFSAGASATSSDRTNDQVNHRTAAGDKADAWTAGLKYDANNIY 253
QY 227 LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTSKAKDVEGI 286
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 LATMYSETRNMTPEGD---SDYAVANKTQNFVETAGYQDFGLRPAVSLMSKGRDLHAA 310
QY 267 GDVD 290
Db 311 GGAD 314

RESULT 6
US-08-096-182A-4
; Sequence 4, Application US/08096182A
; Patent No. 5439808
; GENERAL INFORMATION:
; APPLICANT: Blake, Milan S.
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Huilin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Method for the High Level Expression,
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,182A
; FILING DATE: 23-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0060000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-096-182A-4
Query Match 8.7%; Score 157; DB 1; Length 342;
Best Local Similarity 24.7%; Pred. No. 8.5e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
QY 11 VDLYKAVGLHYFSKNGENSYGGNGDMTY-----ARLGFKEGTQINSDLTGYGQ 61
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 VTLYGTIRAGVEVSRYKADAGTYKAOGKSKTATQIADFGSKIGFGQEDLGNMKAIWQL 62
QY 62 EYNFGNNSGADAGTGNKTRLAFLAGLKADYGVDFGRNRYGVVYDALGYTDMLPFEG 119
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 E---QKASITAGTNSGMGN--RQSFTGLK---GGFTVVRAGNLNTVLKDSGDNVNNAWESG 113
QY 120 GDTATSDOFFEFCVRGVGATYRNSNFFGLVGLNFAVQYLGK---NERDTARRSNGDGV 174
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 SNT--EDVLGLTIGTIGVESREISVRYDSPVAFGSGSVQYVPRDNANDVDKTKHTKSSRE 171
QY 175 GGSISYEVEFGIVGAYGA-----ADR--TNLQEAQPLNGKKAQEWATGLKYD 221
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 SYHAGLKVENAGFFGQYAGSFAKADLNTDAERVAVNTANAHVP---KDYQVHRVVAGYD 228
QY 222 ANNIY--LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTSK 279
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 ANDLYSVAGQYEAANK-----NEVGSTGKKKHEQTQVAATAAYRFG-NVTPRVSYAHGF 282
QY 280 AKDVEIGDVLVNY----FEVGATYTFNKNMSTYV 311
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 KAKVNGVD---ANYQYDQVIVGADYDFSKRTSALV 315
RESULT 7
US-08-877-109-4
; Sequence 4, Application US/08877109
; Patent No. 5747287
; GENERAL INFORMATION:
; APPLICANT: Blake, Milan S.
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Huilin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Method for the High Level Expression,
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,109
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,264
FILING DATE: 28-APR-1995
APPLICATION NUMBER: 08/096,182
FILING DATE: 23-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060001
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-877-109-4

Query Match 8.7%; Score 157; DB 1; Length 342;
Best Local Similarity 24.7%; Pred. No. 8.5e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
QY 11 VDLYGRAVLHYFSKNGNSYSGNGDMY-----ARLGFKGETQINSDLTCYQGW 61
DB 3 VTLGTIKAGVEVSRVKDAGTKYKAQGGKSKTATQIADFGSKIGKQEDLGNMKAIWQL 62
QY 62 EYFQGNSEGADAQGNKTRLAFLAGKYADVGSFDYGR--NYGVYDALGYTDMLEPFG 119
DB 63 E---QKASITAGTNSGNGN--RQSFGLK---GGFCTVRAGNLNTLYKDSGDNVNAWESG 113
QY 120 GDTAYSDDFVGVGVGVAIYRNSNFFG--LVQGLNFAVOYLKG---NERDTARRSNGDGV 174
DB 114 SNT--EDVLGLGTIGRVEISRYVDSVPVAFGSGVQVYVPRDNANDVDKYKHTKSSRE 171
QY 175 GGSISYEYEGFVGVGAYGA-----ADR--TNLQEAQPLGNGKKAQWATGLKYD 221
DB 172 SYHAGLYENAGPFGQYAGSFAYADLNTDAERVAVNTANAHPV---KDYQVHRVVAGYD 228
QY 222 ANNIY--LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTKSK 279
DB 229 ANDLYSVAGQYEAAKN-----NEVGSTKGKKEQTOVAATAAYRFG-NVTPRVSYAHGF 282
QY 280 AKDVEGIGDVLVNY-----FEVGATYFFNKNMSTYV 311
DB 283 KAKVNGVKD---ANYQYDQVIVGADYDFSKRTSALV 315

RESULT 8
US-08-798-760-4
Sequence 4, Application US/08/98760
Patent No. 6013267
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huilin L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
PORIN PROTEINS FROM NEISSERIA MENINGITIDIS
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
PORIN PROTEINS FROM NEISSERIA MENINGITIDIS
NUMBER OF INVENTIONS: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,760
FILING DATE: 11-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060002
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-798-760-4

Query Match 8.7%; Score 157; DB 3; Length 342;
Best Local Similarity 24.7%; Pred. No. 8.5e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
QY 11 VDLYGRAVLHYFSKNGNSYSGNGDMY-----ARLGFKGETQINSDLTCYQGW 61
DB 3 VTLGTIKAGVEVSRVKDAGTKYKAQGGKSKTATQIADFGSKIGKQEDLGNMKAIWQL 62
QY 62 EYFQGNSEGADAQGNKTRLAFLAGKYADVGSFDYGR--NYGVYDALGYTDMLEPFG 119
DB 63 E---QKASITAGTNSGNGN--RQSFGLK---GGFCTVRAGNLNTLYKDSGDNVNAWESG 113
QY 120 GDTAYSDDFVGVGVGVAIYRNSNFFG--LVQGLNFAVOYLKG---NERDTARRSNGDGV 174
DB 114 SNT--EDVLGLGTIGRVEISRYVDSVPVAFGSGVQVYVPRDNANDVDKYKHTKSSRE 171
QY 175 GGSISYEYEGFVGVGAYGA-----ADR--TNLQEAQPLGNGKKAQWATGLKYD 221
DB 172 SYHAGLYENAGPFGQYAGSFAYADLNTDAERVAVNTANAHPV---KDYQVHRVVAGYD 228
QY 222 ANNIY--LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTKSK 279
DB 229 ANDLYSVAGQYEAAKN-----NEVGSTKGKKEQTOVAATAAYRFG-NVTPRVSYAHGF 282
QY 280 AKDVEGIGDVLVNY-----FEVGATYFFNKNMSTYV 311
DB 283 KAKVNGVKD---ANYQYDQVIVGADYDFSKRTSALV 315

RESULT 9
PCT-US94-08327-4
Sequence 4, Application PC/TUS9408327
GENERAL INFORMATION:
APPLICANT: The Rockefeller University
APPLICANT: 1230 York Avenue
APPLICANT: New York, New York 10021
APPLICANT: United States of America
APPLICANT: 12103 Indian Creek Court
APPLICANT: Beltsville, Maryland 20705
APPLICANT: United States of America
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huilin L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.

;; TITLE OF INVENTION: Method for the High Level
;; TITLE OF INVENTION: Expression,
;; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
;; TITLE OF INVENTION: Group B
;; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1100 New York Ave., Suite 600
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3934
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/08327
;; FILING DATE: Herewith
;; CLASSIFICATION:
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/096,182
;; FILING DATE: 23 July 1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Esmond, Robert W.
;; REGISTRATION NUMBER: 32,893
;; REFERENCE/DOCKET NUMBER: 1438.006PC00
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 342 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US94-08327-4

Query Match 8.7%; Score 157; DB 5; Length 342;
Best Local Similarity 24.7%; Pred. No. 8.5e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
Qy 11 VDLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKGTQINSDLTGYQGW 61
Db 3 VTLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKGTQINSDLTGYQGW 62
Qy 62 EYFQGNNSGADAGTGNKRLAFAGLKYADVGSFDYGR--NYGVVYDALGYTDMLEPFG 119
Db 63 E---QKASIAGTNSGWN--RQSFGLK-----GGFTVRAGNLNTVLKDSGDNVNAWESG 113
Qy 120 GDTAYSDFFVGRGVGVATYRNSNFFG--LVDGLNFAYQYLK-----NERDTARRSNGDV 174
Db 114 SNT--EDVLGLGTIGRVESREISRYDSPVAGFSGVQYVPRDNANDVDYKHKTKSRE 171
Qy 175 GGSISYEYEGFIVGAYGA-----ADR--TNLQEAQPLGNGKAEQWATGLKYD 221
Db 172 SYHAGLKYENAGFFQYAGSFAKYADLNTDAERVAVNTANAHV---RDYQVHRVYAGYD 228
Qy 222 ANNIY--LAANYGETRNATPTTNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIATYKSK 279
Db 229 ANDLVSVAGQYEAANK-----NEVGSTGKKHKEQTQVAATAAYRFG-NVTPRVSYAHGF 282
Qy 280 AKDEGIGDVLVNY-----FEVGATYYFNKNMSTVV 311
Db 283 KAKVNGVKD---ANYQYDQIVGADYDFSKRTSALV 315

RESULT 10
US-08-096-182A-6
; Sequence 6, Application US/08096182A
; Patent No. 5439808

;; GENERAL INFORMATION:
;; APPLICANT: Blake, Milan S.
;; APPLICANT: Tai, Joseph Y.
;; APPLICANT: Qi, Huilin L.
;; APPLICANT: Liang, Shu-Mei
;; APPLICANT: Hronowski, Lucjan J.J.
;; APPLICANT: Pullen, Jeffrey K.
;; TITLE OF INVENTION: Method for the High Level Expression,
;; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
;; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1100 New York Ave., Suite 600
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/096,182A
;; FILING DATE: 23-JUL-1993
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Esmond, Robert W.
;; REGISTRATION NUMBER: 32,893
;; REFERENCE/DOCKET NUMBER: 1438.0060000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 363 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-096-182A-6

Query Match 8.7%; Score 157; DB 1; Length 363;
Best Local Similarity 24.7%; Pred. No. 9.2e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
Qy 11 VDLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKGTQINSDLTGYQGW 61
Db 24 VTLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKGTQINSDLTGYQGW 83
Qy 62 EYFQGNNSGADAGTGNKRLAFAGLKYADVGSFDYGR--NYGVVYDALGYTDMLEPFG 119
Db 84 E---QKASIAGTNSGWN--RQSFGLK-----GGFTVRAGNLNTVLKDSGDNVNAWESG 134
Qy 120 GDTAYSDFFVGRGVGVATYRNSNFFG--LVDGLNFAYQYLK-----NERDTARRSNGDV 174
Db 135 SNT--EDVLGLGTIGRVESREISRYDSPVAGFSGVQYVPRDNANDVDYKHKTKSRE 192
Qy 175 GGSISYEYEGFIVGAYGA-----ADR--TNLQEAQPLGNGKAEQWATGLKYD 221
Db 193 SYHAGLKYENAGFFQYAGSFAKYADLNTDAERVAVNTANAHV---RDYQVHRVYAGYD 249
Qy 222 ANNIY--LAANYGETRNATPTTNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIATYKSK 279
Db 250 ANDLVSVAGQYEAANK-----NEVGSTGKKHKEQTQVAATAAYRFG-NVTPRVSYAHGF 303
Qy 280 AKDEGIGDVLVNY-----FEVGATYYFNKNMSTVV 311
Db 304 KAKVNGVKD---ANYQYDQIVGADYDFSKRTSALV 336

RESULT 11
US-08-877-109-6

; Sequence 6, Application US/08877109

; Patent No. 5747287

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Qi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA: US/08/877,109

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

; FILING DATE: 28-APR-1995

Db 250 ANDLYSVAGQYEAANK-----NEVGSTKGKKHEQTQVAAATAAYRFG-NVTPRVSYAHGF 303
QY 280 AKDVEGIGDVLVNY-----FEVGATYYFNKNMSTYV 311
Db 304 KAKVNGVD---ANYQYDQVIGADYDFSKRTSALV 336

RESULT 13
PCT-US94-08327-6
; Sequence 6, Application PC/TUS9408327
; GENERAL INFORMATION:
; APPLICANT: The Rockefeller University
; APPLICANT: 1230 York Avenue
; APPLICANT: New York, New York 10021
; APPLICANT: United States of America
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, Maryland 20705
; APPLICANT: United States of America
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Huilin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Expression, Method for the High Level
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
; TITLE OF INVENTION: Group B
; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08327
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: US 08/096,182
; FILING DATE: 23 July 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 1438.006PC00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08327-6

Query Match 8.7%; Score 157; DB 5; Length 363;
Best Local Similarity 24.7%; Pred. No. 9.2e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLTGKAVCLHYFSGKNGENSYGGNGDMTY-----ARLGFKGETQINSDLTGYGQW 61
Db 24 VTLTGITKAGVEVSVDAGYTKAQGGKSKTQTQIADFGSKIGFKGQEDLGNMKAIWQL 83
QY 62 EYNFGQNNSEGADQGTGNKTLAFAGLKYADVGSFDYGR--NYGVVYDALGYTDLPEFG 119

Db 84 E---OKASIAGTNSGWN--RQSFITGLK-----GGFGTVRAGNLNTVLKDSGDNVANWESG 134
QY 120 GDTAYSDDEFFVGRGVGVATYRNSNEFG--LVDGLNFAYQYLKG---NERDTARRSNGDGV 174
Db 135 SNT--EDVLGLGTIGRVESREISVRYDSPVFAGFSGSVQYVPRDNANDVDKVKHKSSRE 192
QY 175 GGSISYEYEGFGIVGAYGA-----ADR--TNLQEAQPLNGKKAEQWATGLKYD 221
Db 193 SYHAGLKYENAGFTGQYAGSFAKYADLNTDAERVAVNTANAHPV---KDYQVHRVVAGVD 249
QY 222 ANNIY--LAANYGETRNATPTINKETNTSGFANKTQDVLVLAQYQDFGLRLRSIATYTSK 279
Db 250 ANDLYSVAGQYEAANK-----NEVGSTKGKKHEQTQVAAATAAYRFG-NVTPRVSYAHGF 303
QY 280 AKDVEGIGDVLVNY-----FEVGATYYFNKNMSTYV 311
Db 304 KAKVNGVD---ANYQYDQVIGADYDFSKRTSALV 336

RESULT 14
US-08-096-182A-2
; Sequence 2, Application US/08096182A
; Patent No. 5439808
; GENERAL INFORMATION:
; APPLICANT: Blake, Milan S.
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Huilin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Method for the High Level Expression,
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,182A
; FILING DATE: 23-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 32,893
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-182A-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;
Best Local Similarity 25.1%; Pred. No. 5.7e-06;
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

QY 35 NGDMTY-----ARLGFKGETQINSDLTGYGOWEYFQGNNSSEGADAQTGNKTRL 83
Db 19 NGQVTEVTATGIVDLGSKIGFKGQEDLGNLKAIWQVE---OKASIAGTSGWGN--RQ 73

QY 84 AFAGLKADYVGSFSDYGRNYGVVYDALGYTDMLEPFGD-----TAYSDFFVGRVGG---- 135
 Db 74 SFGLK-GGFGKLRVGRNLNVLKDT-----GDNPWDSKSDYLGYNKIAEPAR 121
 QY 136 --VATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISYEYEGFIVGAYGA 193
 Db 122 LISVRYDSPFAGLSGSVQYAL-----NDNAGRHNSESYHAGFNKNGGFFV--QYGG 172
 QY 194 ADRTNLQEAQPLGNK-KAOWATGLKYDANNIYLAANYGETRNPATITNKFTNTSGFAN 252
 Db 173 AYKRHHQVQEGLNIEKYQIHLVSG--YDNDALY-ASVAVQOQDA-----KLTDAASHN 224
 QY 253 KTQDVLVLAQYQDFGLRPSIATYKSKAKDVEG-IGDVLVNYFE---VGTATYFKNMS 308
 Db 225 SQTEVAATLAYRFG-NVTPRVSY-----AHGFKGLVDDADIGNEYDQVWVGAEDFSKRTS 279
 QY 309 TYV 311
 Db 280 ALV 282

RESULT 15

US-08-877-109-2

; Sequence 2, Application US/08877109

; Patent No. 5747287

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Oi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/877,109

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/431,264

; FILING DATE: 28-APR-1995

; APPLICATION NUMBER: 08/096,182

; FILING DATE: 23-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 309 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-877-109-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;

Best Local Similarity 25.1%; Pred No. 5.7e-06;
 Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;
 QY 35 NGDMTY-----ARLFGKETQINSDLTCYGQWEYNFGNNSGADAQTKNTRL 83
 Db 19 NGQVTEVTTATGIVDLGSKIGFKGQEDLGNGLKAIWQVE---QKASITAGTDSGNGN--RQ 73
 QY 84 AFAGLKADYVGSFSDYGRNYGVVYDALGYTDMLEPFGD-----TAYSDFFVGRVGG---- 135
 Db 74 SFGLK-GGFGKLRVGRNLNVLKDT-----GDNPWDSKSDYLGYNKIAEPAR 121
 QY 136 --VATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISYEYEGFIVGAYGA 193
 Db 122 LISVRYDSPFAGLSGSVQYAL-----NDNAGRHNSESYHAGFNKNGGFFV--QYGG 172
 QY 194 ADRTNLQEAQPLGNK-KAOWATGLKYDANNIYLAANYGETRNPATITNKFTNTSGFAN 252
 Db 173 AYKRHHQVQEGLNIEKYQIHLVSG--YDNDALY-ASVAVQOQDA-----KLTDAASHN 224
 QY 253 KTQDVLVLAQYQDFGLRPSIATYKSKAKDVEG-IGDVLVNYFE---VGTATYFKNMS 308
 Db 225 SQTEVAATLAYRFG-NVTPRVSY-----AHGFKGLVDDADIGNEYDQVWVGAEDFSKRTS 279
 QY 309 TYV 311
 Db 280 ALV 282

Search completed: December 18, 2002, 16:23:14

Job time : 8.85111 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:21:10 ; Search time 4.71067 Seconds
(without alignments)
1213.493 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDKNKVDLYGKAVGL.....NKLGVSDDTVAVGIVYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	925.5	51.2	322	10	US-09-905-176-24
2	113	6.2	467	10	US-09-912-020-381
3	112	6.2	207	9	US-10-029-180-70
4	110	6.1	1426	10	US-09-912-020-340
5	104.5	5.8	708	8	US-08-834-666A-2
6	104	5.7	848	10	US-09-841-132-192
7	104	5.7	1530	10	US-09-841-132-178
8	103	5.7	691	8	US-08-834-666A-12
9	103	5.7	691	8	US-08-834-666A-22
10	102	5.6	711	8	US-08-834-666A-8
11	101	5.6	1325	10	US-09-741-669-304
12	100	5.5	1531	12	US-10-007-693-98
13	99	5.5	733	8	US-08-834-666A-4
14	98.5	5.4	597	10	US-09-793-306-146
15	96.5	5.3	440	10	US-09-815-242-5131
16	96.5	5.3	943	9	US-09-996-634-131
17	95.5	5.3	925	9	US-09-924-097-14
18	95	5.3	745	8	US-08-834-666A-6
19	93.5	5.2	909	10	US-09-332-226-4

20	93	5.1	672	8	US-08-834-666A-10	Sequence 10, Appl
21	92.5	5.1	595	10	US-09-881-752A-166	Sequence 166, App
22	92.5	5.1	1536	9	US-10-092-880-2	Sequence 2, Appl1
23	92	5.1	1377	10	US-09-815-242-10384	Sequence 10384, A
24	91.5	5.1	700	10	US-09-881-752A-180	Sequence 180, App
25	90.5	5.0	1430	10	US-09-740-274-6	Sequence 6, Appl1
26	90	5.0	272	10	US-09-119-900-15	Sequence 15, Appl
27	90	5.0	934	8	US-08-837-459-19	Sequence 19, Appl
28	90	5.0	1290	10	US-09-881-752A-138	Sequence 138, App
29	89.5	4.9	815	10	US-09-815-242-5106	Sequence 5106, Ap
30	89	4.9	410	10	US-09-905-176-26	Sequence 26, Appl
31	89	4.9	715	9	US-09-870-759-86	Sequence 86, Appl
32	89	4.9	878	10	US-09-912-020-364	Sequence 364, App
33	89	4.9	2122	9	US-09-813-214A-9	Sequence 9, Appl1
34	88.5	4.9	470	10	US-09-881-752A-174	Sequence 174, App
35	88	4.9	1152	10	US-09-945-265-4	Sequence 4, Appl1
36	88	4.9	1153	10	US-09-350-259-3	Sequence 3, Appl1
37	88	4.9	1974	9	US-09-895-913A-12	Sequence 12, Appl
38	87.5	4.8	839	9	US-10-023-437-23	Sequence 23, Appl
39	87	4.8	271	10	US-09-119-900-8	Sequence 8, Appl1
40	87	4.8	518	9	US-09-976-297-2	Sequence 2, Appl1
41	87	4.8	793	10	US-09-881-752A-362	Sequence 362, App
42	87	4.8	2353	10	US-09-797-862-33	Sequence 33, Appl
43	86	4.8	1600	9	US-10-092-880-10	Sequence 10, Appl
44	85.5	4.7	285	9	US-09-895-913A-138	Sequence 138, App
45	85	4.7	565	9	US-10-108-605-299	Sequence 299, App

ALIGNMENTS

RESULT 1
US-09-905-176-24
; Sequence 24, Application US/09905176
; Patent No. US20020150906A1
; GENERAL INFORMATION:
; APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
; APPLICANT: Debe, Derek A.
; TITLE OF INVENTION: METHOD FOR DETERMINING THREE-DIMENSIONAL PROTEIN STRUCTURE FRO
; FILE REFERENCE: 265/297
; CURRENT APPLICATION NUMBER: US/09/905,176
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US 60/218,016
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 24
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-905-176-24

Query Match	51.2%	Score	925.5;	DB 10;	Length	322;	
Best Local Similarity	57.3%	Pred. NO.	1.1e-76;				
Matches	192;	Conservative	35;	Mismatches	81;	Gaps	8;
QY	10	KVDLYGKAVGLHYFSKXGNGENSYCGNGDMTYARLGFKEGTQINSDLTGYGQWYVNFQGN	69				
Db	1	KLDLYGKIDGLHYFS-----DDRDVGDQTYMRGLGVKGTQINDQLTGYGQWYVNVQANN	55				
QY	70	SEGADATGKNKTRLAFLAGLYADVGSFDYGRNRYGVVYDALGYDMLPEFGDGTAYSDDFF	129				
Db	56	TESSSDQAW--TRLAFLAGLFGDAGSPDYGRNRYGVVYDVTSWTDVLPEFGDGTYSQNF	113				
QY	130	VGRVGVVATYRNSFFGLVDGLNFAVQYLGKN-----ERDT-----ARRNSGQVGGIS	179				
Db	114	QSRANGVATYRNSDFFGLV-GLNFAQYQKNGSVSGEGATNNGRGALKONGDGFQTSVT	172				
QY	180	YE-YEGFVIGVAYCAADRTNLQEAQPLNGCKKAEQATGLKYDANNIYLAANTYLAANTY	238				
Db	173	YDIFPDGISAGFAYANSKRDTDDQNLGLGEGDHAETTYTGLKYDANNIYLAANTYLAANTY	232				

```
QY 239 PITNFTWTSFANKTQDVLLVAQYQDFGLRPSYATYKSKAKDVEGIDVDLVNFEVG 298
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 233 R-----AGSLGFANKAFNEAAQYQDFGLRPSVAYLQSKGKDLNGYGDQDILKYVDVG 287
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
QY 299 ATYFENKKNSTVVDYIINQIDSDN---KLGVGSD 330
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 288 ATYFENKKNSTVVDYIINQIDSDN---KLGVGSD 332
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

RESULT 2
US-09-912-020-381
; Sequence 381, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 381
; LENGTH: 467
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-381

Query Match 6.2%; Score 113; DB 10; Length 467;
Best Local Similarity 22.8%; Pred. No. 0.0079;
Matches 79; Conservative 31; Mismatches 114; Indels 122; Gaps 18;

QY 22 YFSKNGNSYGGNDMTYARLGFKEGTQINSDLFGYQWMEYFQGNSE---GADAQTG 78
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 213 YF---GCFNGNDGTINY-----DQDVNGIMVG---DTKIDGNNAKIVGAAA--- 255
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 79 NKTRLAFLAGLVADGSDYGRNYGVVVDALGYTMDLPEFGDGTAYSDDFVGRVGGVAT 138
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 256 -----GFAKGDMDNRSGQVDQ-----DSQTAY-----IYSSAH 283
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 139 YRNSNFFGLVDLNPVAVQYLGKNERDTRRSNGDGVGSGSISYEYEGFGI----- 187
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 284 FANNVF---VDG---SLSYSHENNDSLNTSMNGTVVDGSTNSDANGFLKAGYDFKLGA 337
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 188 ---VGAYGADRT-NLQEAQPLGNGKAKQWATGLKYDANNIYLAANYGETRNPATITNKF 244
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 338 GYVTPYGVSGVGLFQSGDDYQLSDNMKVD---GQSYDSMRVELGVDA GT-----F 384
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 245 TWTSGFANKTQDVLLVAQYQDFGLRP--STAYTSKAKDVEGIDVDLVNFE----- 296
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 395 T-----YSEDQALTPFKLAYVVDSDNNNDVNGDSIDNGTEGSAVRV 427
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 297 -VGATPYFENKKNSTVVDYIINQIDSDNKLGVGSDD---TVAVGIYV 338
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 428 GLGTQFSFTKNFSATD-----ANYLGGDQVDQDMSANVGKY 465
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

RESULT 3
US-10-029-180-70
; Sequence 70, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Call, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffry C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: NO. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-70

Query Match 6.2%; Score 112; DB 9; Length 207;
Best Local Similarity 24.8%; Pred. No. 0.0035;
Matches 51; Conservative 24; Mismatches 99; Indels 32; Gaps 9;

QY 4 YNKDGNKVDLYGKAVGLHYFSGKNGNSYGGNDMTYARLGFKEGTQINSDLTGQWHEY 63
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 7 YNQGYYGQGGYGGQ---GYGPGYGGYGGYGGQPG---YGGQGHQDQGGYGGYGGYGGY 61
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 64 NFGGNNS-----EGADAQTGNKTRLAFLAGLVAD---VGSFDTGRNYGVVVDALGY 111
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 62 GQGGSSDYAGQQHQGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGY 120
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 112 TMDLPEFGDGTAYSDDFVGRVGGVATYR-NSNFFGLVDG--LNFVAVQYLGKNERDTRR 168
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 121 PDEAQD--GERGIAGALAGAAAGGAGKHVNHGFLGTGGTIGGALIGSIAEDAVKKHRNSDQ 178
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

QY 169 S-----NGDGVGSGSISYEYEGF 185
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 179 SPPQYGGPPPSNSGSGSGSMMDQLGGF 204
; : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||

RESULT 4
US-09-912-020-340
; Sequence 340, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
```

US-09-912-020-340

Query Match 6.18; Score 110; DB 10; Length 1426;
Best Local Similarity 21.88; Pred. No. 0.06;
Matches 88; Conservative 46; Mismatches 143; Indels 126; Gaps 19;

QY 2 EYKNGKNDLYKAVGL-----HYFSKG-----NGE-----NS 31
DB KFIYNDGNQLTAVVSPDGLSRRYDEPGRVLSYTSRSGEIVRYDDAHSELPAITTTDA 528
QY 32 YGGNGDMYARLGFQETQINSLDTGY-----GQM-----EYNFOGNN 69
DB TGSTROMTWSRY---GQLLAFTDCGQYRYEYDRFGOMTAVHREBEGISLYRRYDNRGL 585
QY 70 SEGADAQTKNTRLAFLAGKYADVSDPY-----GRNVGVYDALGYDMLPERGGDTAY 124
DB TSVKDAQ-GRETR-----YEYNAAGDITAVITPDGNRSETQYDAMKAVSTTQ-GGLTRS 638
QY 125 SDDFFVGRVGVATYRNSNFFGLVDGLNFAVOYLKGNKRDTPARRSNGDVGSGISYEYEG 184
DB MEYDAAGRVISLTNENGSHSVFSDALDRLVQOGGFDGR--TQYHYDLTGKLTQSEDEG 696
QY 185 FGIYGAAGADRTWLQEAQPLGNKKAQ-----WATGLKYDANNIYLAANYG----- 232
DB LVILWYDESDRITHRV---NGEPAEQWQYDGHGWLTDIHLSEGHRAVAVHYGYDDKG 752
QY 233 -----ETRNATPITNKFTNTSGFANK-TQDVLLVNAQY----- 263
DB RLACGECQTVENPERGELLMOHETKHA-----YNEOGLANRVTPDSLPPVWEWTYSGGY 805
QY 264 --QPDFGLRPSIATYTKS--AKDVEGIGDVLVN-YFEVGATY 301
DB LAGMKLGTPVETRDLRHRETVRSFGSMAGSNAAYELTSTY 848

RESULT 5

US-08-834-666A-2
Sequence 2, Application US/08834666A
Patent No. US2002004949A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Lissolo, Ling
APPLICANT: Tomb, Jean-Francois
APPLICANT: Miller, Charles
APPLICANT: Al-Garawi, Amal
TITLE OF INVENTION: 76 Kda Helicobacter Polypeptides and
TITLE OF INVENTION: Corresponding Polynucleotide Molecules
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834, 666A
FILING DATE: 01-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,175
REFERENCE/DOCKET NUMBER: 06132/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...19
OTHER INFORMATION:
US-08-834-666A-2

Query Match 5.88; Score 104.5; DB 8; Length 708;
Best Local Similarity 20.88; Pred. No. 0.079;
Matches 80; Conservative 53; Mismatches 145; Indels 107; Gaps 20;

QY 1 AEIYNKDKGNKVDLYGK-----AVGLHYFSKNGENSYGGNGDMTVARLGF-----K 46
DB AEILNQARQVVKNFEKIPAFVSDSLGVCYEVQG--GERRGTNPQOVTSNTWGACAYVKQ 415
QY 47 GETOINSLDTGYGOWEYNFOGNNSEG-ADAQTGNKTRLAFLAGKYADVSGSDYGRNIGV 105
DB TITLNDNSIAHFGTQEQIQ--QAEIADTLVNFPSR-----YSELGN-----TYNSI 461
QY 106 YDALGYDMLPEFGG-----DTAYSDDFVGRVGVAT--YRNSNFFGLVDGLNFAVOYL 158
DB TTALSCKVPNAQSLQNVVSKNNPSPQ-----GIETNYLQNSYNQIQITIN---QEL 511
QY 159 GKNERD-----TARRSNG--DGVGSGISYE-----YEGF----- 185
DB GRNPERKVGIVNSQTNNGAMGIGIQGYKOFFQOKRWGARYYGFDDYNNHAFKSSFFN 571
QY 186 --GIVGAYG-AADR-----TNLQEAQPLGNKKAQEWATGLKYDANNIYLAANYGETRAT 238
DB SASDVWTVYFGADALYNFINDKATNFKLGNKNSLVGLFGGIALAGTSLNSEYVNLATVN 631
QY 239 PITNKFNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVL----- 291
DB NYNAKMNVANF-----QFLFNMGYRMNLARSKRKGSDHAAHQHIGELGLKIPTI 680
QY 292 -VNYFE-VGATYFENKNNSTYVDYI 314
DB NTNYSFPGAEKLYRRLYSVLNV 705

RESULT 6
US-09-841-132-192
Sequence 192, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhatia, Ajay
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Probst, Peter
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 192
LENGTH: 848
TYPE: PRT
ORGANISM: Chlamydia
US-09-841-132-192

Query Match 5.78; Score 104; DB 10; Length 848;
Best Local Similarity 22.68; Pred. No. 0.11;
Matches 72; Conservative 39; Mismatches 103; Indels 104; Gaps 20;

QY 29 ENSYGGNGDMTVARLGFQETQINSLDTGY-----GQWEYNFOGNNSEGADA 75
DB KFIYNDGNQLTAVVSPDGLSRRYDEPGRVLSYTSRSGEIVRYDDAHSELPAITTTDA 528

QY 283 VEGIGDVL-----VNFE-VGATYFKNKMSYVDYI 314
Db 668 HSAHQHMGELGKIPTINTNTYSLTGTLOYRRLYSVLNVV 708

RESULT 11

US-09-741-669-304
; Sequence 304, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 304
; LENGTH: 1325
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-304

Query Match 5.6%; Score 101; DB 10; Length 1325;
Best Local Similarity 21.4%; Pred. No. 0.36;
Matches 75; Conservative 45; Mismatches 133; Indels 98; Gaps 19;

QY 14 YG-KAYGLHYFSK---GNGENSYGGNGDMTYARLFGKGETQIN-----SDLTGHWGWE-- 62
Db 986 YGLKALNIHGQKLTAEHGGAYGATADMS-AKIGEGDLAINTVQVSLNSQNDYQGA 1044
QY 63 -YFQNNSEGADAQTGNKTRIAFLAGLYADVGSFDYGRNYYGVVYDALGYTDMLEPFGGD 121
Db 1045 TYVQMTLRTDADLGNTRNLNLSN-----AAVDLNGSTQTV----- 1083
QY 122 TAYSDFFVGRVGVATYRNSNFFGLVGLNFVAVQYLGKNERDTARRSNGDVGGSISYE 181
Db 1084 -----ETFGQMGSTVLFKEG-----ALTVNKGGISQGGELTGGGNLNTVGGTFLATE 1129
QY 182 YEGFGIVGAYGAAD-----RTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETR 235
Db 1130 ---GLNARYNALTSISPAEVSLDNTQGLGRGNIAANDGLLTIK-----NV-----TGLRL 1176
QY 236 NATPITNKFTNTSGPANKT--QDVLL-----VAQYQDFGLRPSIATYTKSKAKDVEGI 286
Db 1177 NS--ISGK-----GIVSATARTDVELDGDNSRFVQGFQFNIDTGSALSV-----NEQKNL 1222
QY 287 GDVDLVNVEVGATYFKNKMSYVDYIINQIDSDNKLGVG---SDDTVA 333
Db 1223 GDASVIN-----NGLLTISTERSWAMTHSISGSGDVTKLGTGILTLNDSAA 1269

RESULT 12

US-10-007-693-98
; Sequence 98, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 98
; LENGTH: 1531
; TYPE: PRT

; ORGANISM: Chlamydia trachomatis serovar D
US-10-007-693-98

Query Match 5.5%; Score 100; DB 12; Length 1531;
Best Local Similarity 22.0%; Pred. No. 0.53;
Matches 72; Conservative 42; Mismatches 109; Indels 104; Gaps 20;

QY 20 LHVFSKNGENSYGGNGDMTYARLFGKGETQINSDLTGY-----COWEYNFQ 66
Db 1167 IHVTFEIEDTYGHMGDMSEAKIQ-DGLVLSWNTGTGRDLPQKAGALVFNALWE---- 1221
QY 67 GNNSEGADAQTGNKTRIAFLAGLYADVGSFDYGRN-YGVVYDALGYTDMLE----- 117
Db 1222 ----EGAVLSALKNAR--FAHNLTQORMEFDYSTNVWGFAFG--GRTLSAENLVAIDGY 1273
QY 118 ---FGG-----DTAYSDDDFFVGRVGVATYRNSNFFGLVGLNFVAVQYLGKNERDTARRS 169
Db 1274 KGAYGGASAGVDIQLMEDEVLG-VSGAA-----FLGKMDSQKF-----DAEVSRLK- 1317
QY 170 NGDVGGSISYEGFGIVGAYGAADRTNLQEAQPLGN-----GKKAQWAT-GLK 219
Db 1318 ---GVVGSV---YTGF-LAGSMFFKQYSLGETQ---NDMKTRYGVLGSSASWTSRGVL 1367
QY 220 YDA-----NNIYLAANYGETRNATPITNKFTNTSGFANKTQDVLLV 260
Db 1368 ADALVEYRSLVGPVRPTFVALHFNPPYEVSYASMK-----FPGFTGQGREARSFEDASL- 1421
QY 261 AQYQDFGLRPSIATYTKSKAKDVEGIG 287
Db 1422 TMTITPLGKMFELAFIKGFSEVNSLG 1448

RESULT 13

US-08-834-666A-4
; Sequence 4, Application US/08834666A
; Patent No. US20020044949A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
; Corresponding Polynucleotide Molecules
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,666A
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 amino acids
; TYPE: amino acid

[illegible]

```
QY 130 -----VGR-----VGVVA-----TYRNSNFFGLVD----- 149
Db 166 YDIADWVNDNVGTGTOASSVVGGMATLSGVFSENNDTDGDSVKRYNLRGVFAPLHEPG 225
QY 150 -----CLNEFAVOYLGNERNDTARR-----SNGDG--VGGSTSYE--- 181
Db 226 NVVHLGLQYAYRDLEDSAYDTRIRPRMGMRGVSTNGGNDAGSNGNRGLFGGSSAVEGLWK 285
QY 182 -----YEGFIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETR 235
Db 286 DDSVMGLEGAWALGAFSA-----QAEYLRRTVKAERDREDLK--ASGYAQLAY--TL 334
QY 236 NATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRFSIATYTKSKAD-----VEGIG 287
Db 335 TGEPRLYKLDGAKFDTIKPENKEIGAWELF-----YRYDSIKVEDDNIIVDSATREVG 387
QY 288 DVDLVNYPFVGATYYFNKNMSTVVDYIINOIDS-DNKLGVGSDDTVAVGIVYQF 340
Db 388 DAKGKTH-TLCVNWYANEAVKVSANYVKAKTDKISNANGDDSGDGLVMRLQYVF 440
```

Search completed: December 18, 2002, 16:39:31
Job time : 9.71067 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 10.0756 Seconds
(without alignments)
3253.588 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDGKVDLYCKAVGL.....NKLGVSDTDVAVGIYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues 283224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1805	99.8	362	1 MMSCF	outer membrane por
2	1789	98.9	362	2 D90755	outer membrane por
3	1789	98.9	362	2 D85619	outer membrane por
4	1113	61.5	351	2 D90662	outer membrane por
5	1113	61.5	351	2 B85513	outer membrane por
6	1110	61.4	351	1 MMSCE	outer membrane por
7	1082	59.8	371	2 AH0147	outer membrane por
8	1070	59.1	351	2 S25520	outer membrane por
9	1058	58.5	350	2 S25525	outer membrane por
10	1058	58.5	350	2 AF0543	outer membrane por
11	1051	58.1	365	1 MMBPP2	outer membrane por
12	1051	58.1	367	1 MMECP2	outer membrane por
13	1044	57.7	350	2 S36606	outer membrane por
14	1042	57.6	365	1 MMECNC	outer membrane por
15	1035	57.2	377	2 D64888	outer membrane por
16	1033	57.1	367	2 H91016	outer membrane por
17	1033	57.1	367	2 B85861	outer membrane por
18	1031.5	57.0	363	2 AE0616	outer membrane pro
19	1030	56.9	363	2 S51104	outer membrane pro
20	1029.5	56.9	363	2 S43159	outer membrane pro
21	1027	56.8	383	2 AE0690	outer membrane pro
22	1023.5	56.6	378	1 MMEBPC	outer membrane pro
23	1023.5	56.6	378	2 AE0789	outer membrane pro
24	1000.5	55.3	378	1 AE9139	outer membrane pro
25	998.5	55.2	366	2 D90888	outer membrane por
26	998.5	55.2	366	2 E85729	putative outer memb
27	992	54.8	360	2 AE0172	probable outer mem
28	976.5	54.0	394	2 S34263	outer membrane por
29	976.5	54.0	394	2 AC0755	outer membrane pro

30	960	53.1	374	2 AH0149	outer membrane pro
31	885	48.9	369	2 S70847	outer membrane por
32	770	42.6	315	2 T46993	hypothetical prote
33	649.5	35.9	255	2 F90878	hypothetical prote
34	635.5	35.1	255	2 D85740	partial probable o
35	573.5	31.7	382	2 G84971	ompF-like porin [I
36	539.5	29.8	214	2 F90966	probable outer mem
37	539.5	29.8	224	2 F85814	probable outer mem
38	486	26.9	171	2 H64960	outer membrane por
39	423	23.4	191	2 G90966	hypothetical prote
40	415	22.9	191	2 G85814	probable outer mem
41	372.5	20.6	123	2 G90878	hypothetical prote
42	372.5	20.6	123	2 C85740	partial probable o
43	351	19.4	134	2 B64961	outer membrane por
44	236.5	13.1	350	2 A82299	outer membrane pro
45	233.5	12.9	96	2 C64900	outer membrane por

ALIGNMENTS

RESULT 1

MMECF

outer membrane porin ompF precursor - Escherichia coli (strain K-12)
N:Alternate names: outer membrane protein 1a; outer membrane protein b
C:Species: Escherichia coli
C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 01-Mar-2002
C:Accession: A93449; A91301; A90314; A25029; H64832; A03431
R:Inokuchi, K.; Mutoh, N.; Matsuyama, S.; Mizushima, S.
Nucleic Acids Res. 10, 6957-6968, 1982
A:Title: Primary structure of the ompF gene that codes for a major outer membrane pro
A:Reference number: A93449; MUID:83090452; PMID:6294623
A:Accession: A93449
A:Molecule type: DNA
A:Residues: 1-362 <INO>
A:Cross-references: GB:J01655; GB:M10311; GB:M10312; MID:g147009; PIDN:AAA24244.1; PI
A:Experimental source: strain K12
R:Mutoh, N.; Inokuchi, K.; Mizushima, S.
FEBS Lett. 137, 171-174, 1982
A:Title: Amino acid sequence of the signal peptide of OmpF, a major outer membrane pr
A:Reference number: A91301; MUID:82139379; PMID:7037455
A:Accession: A91301
A:Molecule type: DNA
A:Residues: 1-37 <MUT>
R:Chen, R.; Kramer, C.; Schmidmayr, W.; Chen-Schmeisser, U.; Henning, U.
Biochem. J. 203, 33-43, 1982
A:Title: Primary structure of major outer-membrane protein I (ompF protein, porin) of
A:Reference number: A90314; MUID:82256494; PMID:7049161
A:Accession: A90314
A:Molecule type: protein
A:Residues: 23-87, 'E', 89-138, 'G', 140-283, 'L', 285-362 <CHE>
R:Nogami, T.; Mizuno, T.; Mizushima, S.
J. Bacteriol. 164, 797-801, 1985
A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous r
A:Reference number: A91809; MUID:86033642; PMID:2997131
A:Accession: A25029
A:Molecule type: DNA
A:Residues: 33-63 <NOG>
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64832
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-362 <BLAT>
A:Cross-references: GB:AE0000195; GB:U00096; MID:g1787156; PIDN:AAC74015.1; PID:g17871
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ompF; toLF; cmlB; coa; cry
A:Map position: 21 min
C:Complex: homotrimer

C:Function: <POR>
A:Description: Forms aqueous channels that facilitate diffusion of small hydrophilic mole
C:Function: <REC>
A:Description: receptor for bacteriophage T2
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-72/Domain: signal sequence #status predicted <SIG>
F:73-362/Product: outer membrane porin ompF #status experimental <MAT>

Query Match 99.8%; Score 1805; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 9.3e-123; Indels 0; Gaps 0;
Matches 340; Conservative 0; Mismatches 0;

QY 1 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60
|||||
DB 23 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82
|||||

QY 61 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 120
|||||
DB 83 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 142
|||||

QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180
|||||
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 202
|||||

QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
|||||
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 262
|||||

QY 241 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 300
|||||
DB 263 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 322
|||||

QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
|||||
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362
|||||

RESULT 2
D90755
outer membrane protein la ECs1012 [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90755
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90755
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034435.1; PID:q13360471; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs1012
C:Superfamily: outer membrane protein phoE

Query Match 98.9%; Score 1789; DB 2; Length 362;
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Indels 0; Gaps 0;
Matches 338; Conservative 0; Mismatches 2;

QY 1 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60
|||||
DB 23 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82
|||||

QY 61 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 120
|||||
DB 83 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 142
|||||

QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180
|||||
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 202
|||||

QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
|||||
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 262
|||||

QY 241 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 300
|||||
DB 263 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 322
|||||

QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
|||||
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362
|||||

RESULT 2
D90755
outer membrane protein la ECs1012 [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90755
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90755
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034435.1; PID:q13360471; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs1012
C:Superfamily: outer membrane protein phoE

Query Match 98.9%; Score 1789; DB 2; Length 362;
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Indels 0; Gaps 0;
Matches 338; Conservative 0; Mismatches 2;

QY 1 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60
|||||
DB 23 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82
|||||

QY 61 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 120
|||||
DB 83 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 142
|||||

QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180
|||||

DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 202
QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
|||||
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 262
|||||
QY 241 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 300
|||||
DB 263 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 322
|||||

QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
|||||
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362
|||||

RESULT 3
B85619
outer membrane protein la (Ia,b,F) [imported] - Escherichia coli (strain O157:H7, sub
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85619
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
illet, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <STO>
A:Cross-references: GB:AE005174; NID:g12514104; PIDN:AA055414.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ompF
C:Superfamily: outer membrane protein phoE

Query Match 98.9%; Score 1789; DB 2; Length 362;
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Indels 0; Gaps 0;
Matches 338; Conservative 0; Mismatches 2;

QY 1 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60
|||||
DB 23 AEIYNKDGKVDLYKAVGLHYFSGKNGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82
|||||

QY 61 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 120
|||||
DB 83 WEYNFQGNNSGADATQGNKTRAFAGLYADVGSFDYGRNYYVVDALGYDMLPEFGG 142
|||||

QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180
|||||
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 202
|||||

QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 240
|||||
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRATPI 262
|||||

QY 241 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 300
|||||
DB 263 TNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYKSKADVEGIDVDLNYFEVGAT 322
|||||

QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
|||||
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362
|||||

RESULT 4
D90662
outer membrane pore protein PhoE ECs0268 [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90662
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

Db 77 WEAEFAGNKAESDTAQ--QKTRLAFLAGLKYKDLGSGFDYGRNLGALYDVEAWTMDFPPEGG 134

QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDRTARRNSNGDVGGSIS 179

Db 135 DSSAQTDNEWTKRASGLATYRTDFFGVIDGLNLTLYQKGNERNRVYKKQNGDGFGLT 194

QY 180 YEYEG--FGIVGAYGAADRNLQBAQPLNGKKAEQWATGLKYDANNIYLAANYGETRNA 237

Db 195 YDFGSDFAISGAYTNSDRTEQNLQSRGTGKRAEAWATGLKYDANNIYLATFYSETRKM 254

QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTKSKAKDVEGIDGVDLVNYFEV 297

Db 255 TPI-----TGGFANKTQNFEAQYQDFGLRPSLGSYLSKGDIEGIDEDLVNYIDV 308

QY 298 GATYFYNKKNSTVYDYTIINQIDSNKLGVSDDTVAVGIVYQF 340

Db 309 GATYFYNKKNSAFYVDYKINQLNDSNKLINNDIDIVAGWTYQF 351

RESULT 6

MMECPE

outer membrane porin phoE precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 01-Mar-2002

C:Accession: A03432; B64749

R:Overbeek, N.; Bergmans, H.; van Mansfeld, F.; Lugtenberg, B.

J. Mol. Biol. 163, 513-532, 1983

A:Title: Complete nucleotide sequence of phoE, the structural gene for the phosphate A:Reference number: A92893; MUID:83189086; PMID:6341601

A:Accession: A03432

A:Molecule type: DNA

A:Residues: 1-351 <OVE>

A:Cross-references: GB:D83536; NID:g4902908; PIDN:BAA7910.1; PID:g4902976

A:Experimental source: strain K12

A:Note: this is one of the proteins induced when the E. coli cells are grown under phoE d compounds, and some other negatively charged solutes

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; .A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64749

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-351 <BLAT>

A:Cross-references: GB:AE000132; GB:U00096; NID:g2367098; PIDN:NAAC73345.1; PID:gl7864

A:Experimental source: strain K-12, substrain MGL655

C:Genetics:

A:Gene: phoE

A:Map position: 6 min

C:Complex: homotrimer

C:Function:

A:Description: allows passive diffusion of small hydrophilic molecules with molecular C:Superfamily: outer membrane protein limitation

C:Keywords: homotrimer; membrane protein; porin

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-351/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 61.4%; Score 1110; DB 1; Length 351;

Best Local Similarity 62.1%; Pred. No. 1.le-72;

Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNGDWTYARLGFKGETQINSLDTGCG 60

Db 22 AEIYNKDGKLDVYGVKVAHYMS-----DNASKDGQDSVIRFGFKGETQINDLTGYGR 76

QY 61 WEYNFQGNNSGDAQGCKTKRLAFLAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 120

Db 77 WEAEFAGNKAESDTAQ--QKTRLAFLAGLKYKDLGSGFDYGRNLGALYDVEAWTMDFPPEGG 134

QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDRTARRNSNGDVGGSIS 179

Db 135 DSSAQTDFNTWRASGLATYRNTDFGVVDGLNLTLYQYQGNENRDYKQNGDGFSTLT 194

QY 180 YEYEG--FGIVGAYGAADRNTLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNA 237

Db 195 YDFGSDFAISGAYTNSDRTEQNLQSRGTGRKRAEAWATGLKYDANNIYLATFYSETRKM 254

QY 238 TPTTNKFTWTSFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVLDLVNFEV 297

Db 255 TPI-----TGGFANKTQNFVAQYQDFGLRPSLGYVLSKGKDFEGIDEDLVNIDV 308

QY 298 GATYFYNKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

Db 309 GATYFYNKMSAFVDYKINQLSDNKLINNDNDIVAVGTYQF 351

RESULT 7

AH0147

outer membrane protein C2 ompC2 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001

C:Accession: AH0147

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0147

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <CUR>

A:Cross-references: GB:AL590842; PIDN:CAC90043.1; PID:g15979264; GSPDB:GN00175

C:Genetics:

A:Gene: ompC2

C:Superfamily: outer membrane protein phoE

Query Match 59.8%; Score 1082; DB 2; Length 371;

Best Local Similarity 59.4%; Pred. No. 1.2e-70;

Matches 212; Conservative 48; Mismatches 73; Indels 24; Gaps 7;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNGDMTYARLKGFTQINSDLTGYGQ 60

Db 22 AEIYNKDGKNDLYGKVKALHFS-----DNTKSDGDKSYVRLGKGVYQITDELSGYQ 76

QY 61 WEYFQGNNSGADQAOTGNKTRLAFAGLADYGVDFGRNYGVVYDALGYTDMLPFEG 120

Db 77 WEYFNAANYAESQEAQ-DNKTRLAFAGLRYNLGSDYGRNYGVLYDIAAWTDMLPFEGN 135

QY 121 DT-AYSDDFEVRGVGVATYRNSNFFGLVDGLNFAVQYLGKN-----ERDTARRSN 170

Db 136 DSYTRTDNFMGTGRTGVATYRNTDFGLVDGLKFLSQYQGNKAGAGETNNGRTDTSKO-N 194

QY 171 GDCVGGSGISYEY-EGFGIVGAYGAADRNTLQEAQPLNGKKAQWATGLKYDANNIYLA 229

Db 195 GDGFGLSSEYIGAGVSGAAYASSNRLTAQKNSTFGKDKADAWTGLKYSNGVYLAA 254

QY 230 NYGETRNATPIT-----NKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDV 283

Db 255 NYAETRNMTPISGTAVINNVSYSVGFANKTQNIELVAQYLFDFGLKPSIAYIQSGKDI 314

QY 284 EGIGDVLVNYFEVATYFYNKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

Db 315 EGIGDVLVYKVIDGATYFYNKMSYVDYKINQLNDNKLKLTNDNVVALGLVYQF 371

RESULT 8

S25520

outer membrane porin phoE precursor - Citrobacter freundii

C:Species: Citrobacter freundii

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999

C:Accession: S25520

R:Spielerings, G.; Ockhuysen, C.; Hofstra, H.; Tommassen, J.

submitted to the EMBL Data Library, August 1992

A:Reference number: S25520

A:Accession: S25520

A:Molecule type: DNA

A:Residues: 1-351 <SPI>

A:Cross-references: EMBL:X68021; NID:g40474; PIDN:CAA48162.1; PID:g40475

A>Note: this is one of the proteins induced when the E. coli cells are grown under ph

d compounds, and some other negatively charged solutes

C:Genetics:

A:Gene: phoE

C:Complex: homotrimer

C:Function:

A:Description: allows passive diffusion of small hydrophilic molecules with molecular

A>Note: induced under phosphate limitation

C:Superfamily: outer membrane protein phoE

C:Keywords: homotrimer; membrane protein; porin

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-351/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 59.1%; Score 1070; DB 2; Length 351;

Best Local Similarity 59.2%; Pred. No. 8.1e-70;

Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNGDMTYARLKGFTQINSDLTGYGQ 60

Db 22 AEIYNKDGKNDLYGKVKAMHYMTDYDSK-----DGDQSYIRLGFKEGTQINDELTYGR 76

QY 61 WEYFQGNNSGADQAOTGNKTRLAFAGLADYGVDFGRNYGVVYDALGYTDMLPFEG 120

Db 77 WEAEFAGNKAESDSNQ--OKTRLAFAGSKLKLGLSFDYGRNLGALYDVAWTDMPFEG 134

QY 121 D-TAYSDDFEVRGVGVATYRNSNFFGLVDGLNFAVQYLGKNERDPTARRSNGDVGSGSIS 179

Db 135 DSSAQTDFNTWRASGLATYRNTDFGVVDGLDLYQYQGNKQNRDVKQNGDGFSTYT 194

QY 180 YEYEG--FGIVGAYGAADRNTLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNA 237

Db 195 YDFGSDFAVSGAYTNSDRTNQNLQTRGTGDKAEAWATGLKYDANDIYATFYSETRNM 254

QY 238 TPTTNKFTWTSFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVLDLVNFEV 297

Db 255 TPI-----SGGFANKTQNFVAQYQDFGLRPSLGYVLSKGKDFEGVNEEDLVNIDV 308

QY 298 GATYFYNKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

Db 309 GATYFYNKMSAFVDYKINQLSDNKLINNDNDIVAVGMYQF 351

RESULT 9

S25525

outer membrane porin phoE precursor - Salmonella typhimurium

N:Alternate names: phosphate-limitation-inducible outer membrane pore protein phoE

C:Species: Salmonella typhimurium

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Aug-1999

C:Accession: S25525; JQ1937

R:Spielerings, G.; Elders, R.; van Hith, B.; Hofstra, H.; Tommassen, J.

submitted to the EMBL Data Library, August 1992

A:Reference number: S25525

A:Accession: S25525

A:Molecule type: DNA

A:Residues: 1-350 <SPI>

A:Cross-references: EMBL:X68023; NID:g47821; PIDN:CAA48164.1; PID:g47822

A:Experimental source: strain LT(2)

R:Spielerings, G.; Elders, R.; van Lith, B.; Hofstra, H.; Tommassen, J.

Gene 122, 45-52, 1992

A:Title: Characterization of the Salmonella typhimurium phoE gene and development of

A:Reference number: JQ1937; MUID:93083994; PMID:1280609

A:Accession: JQ1937

A:Molecule type: DNA

A:Residues: 1-197;199-350 <SP2>

A:Cross-references: EMBL:X68023

A>Note: neither the full codon nor the amino acid are given for residue 198 in Figure

C:Genetics:

A:Gene: phoE

C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-330/Product: outer membrane porin phoE #status predicted <MAT>

Query Match 58.5%; Score 1058; DB 2; Length 350;
Best Local Similarity 58.0%; Pred. No. 5.9e-69;
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYNKDGKVDLYGRAVGLHYFSKNGENSNGGDMYARLGFKEGTQINSDLTGYGQ 60
Db 21 AEIYNKNGKLDYGVKAMHYMSDYSK-----DGDQSYVRFEGFKGTQINDLTGYGR 75
QY 61 WEYNFGNNSGADAQGNKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 120
Db 76 WEAEFASNAKESDSSQ--QKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 133
QY 121 D-TAYSDDFVGRVGVVATYRNFFGLVDGLNFVAQYLGKNERDTRARRSNGDVGGSIS 179
Db 134 DSSAQTDNFMTRKASGLATYRNTRDFFGIVDGLDLTLQYQKNGEDRDVKKQNGDGFCTS 193
QY 180 YEYEG--FGIVGAYGAADRTNLOEAQPLGNGKAEQWATGLKYDANNIYLAANYGETRNA 237
Db 194 YDFGSDFAVSGAYTLSDRTREQLRRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253
QY 238 TPTTNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEV 297
Db 238 TPTTNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEV 297
QY 254 TPV-----SGGFANKTQNFVAIQYQDFGLRPSLGVLGKDGIEGVSDELVNYIDV 307
Db 254 TPV-----SGGFANKTQNFVAIQYQDFGLRPSLGVLGKDGIEGVSDELVNYIDV 307
QY 298 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 340
Db 298 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 340
QY 308 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 350
Db 308 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 350

RESULT 10

AF0543
outer membrane pore protein E precursor STY0365 [imported] - Salmonella enterica subsp.
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF0543
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0543
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-350 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08790.1; PID:g16501606; GSPDB:GN00176
C:Gene: STY0365
C:Superfamily: outer membrane protein phoE

Query Match 58.5%; Score 1058; DB 2; Length 350;
Best Local Similarity 58.0%; Pred. No. 5.9e-69;
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYNKDGKVDLYGRAVGLHYFSKNGENSNGGDMYARLGFKEGTQINSDLTGYGQ 60
Db 21 AEIYNKNGKLDYGVKAMHYMSDYSK-----DGDQSYVRFEGFKGTQINDLTGYGR 75
QY 61 WEYNFGNNSGADAQGNKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 120
Db 76 WEAEFASNAKESDSSQ--QKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 133
QY 121 D-TAYSDDFVGRVGVVATYRNFFGLVDGLNFVAQYLGKNERDTRARRSNGDVGGSIS 179
Db 134 DSSAQTDNFMTRKASGLATYRNTRDFFGIVDGLDLTLQYQKNGEDRDVKKQNGDGFCTS 193

QY 180 YEYEG--FGIVGAYGAADRTNLOEAQPLGNGKAEQWATGLKYDANNIYLAANYGETRNA 237
Db 194 YDFGSDFAVSGAYTLSDRTREQLRRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253
QY 238 TPTTNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEV 297
Db 254 TPV-----SGGFANKTQNFVAIQYQDFGLRPSLGVLGKDGIEGVSDELVNYIDV 307
QY 298 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 340
Db 308 GATYYFNKNNSTVVDYIIQIDSDNKLGVGSDDTVAVGIVYQF 350

RESULT 11

MMBPP2
outer membrane porin lc precursor - phage PA2
C:Species: phage PA2
A:Note: host Escherichia coli
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 02-Jul-1998
C:Accession: D25647
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and nmpc outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457; PMID:3017988
A:Accession: D25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
C:Genetics:
C:Gene: lc
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 58.1%; Score 1051; DB 1; Length 365;
Best Local Similarity 59.8%; Pred. No. 2e-68;

Matches 211; Conservative 35; Mismatches 83; Indels 24; Gaps 7;
QY 1 AEIYNKDGKVDLYGRAVGLHYFSKNGENSNGGDMYARLGFKEGTQINSDLTGYGQ 60
Db 24 AEIYNKDSNKLGLYKVNNAKHYSNDAD-----DGDYTVARLGFKEGTQINDLTGFG 78
QY 61 WEYNFGNNSGADAQGNKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 120
Db 79 WEYEFAGNRAE--SOGSSKOKTRLAFLAGLVADYSGDYGNYGVVYDALGYTDMLPFEFG 137
QY 121 DT-AYSDDFFVGRVGVVATYRNFFGLVDGLNFVAQYLGKNER--DTARRSNGDVG 176
Db 138 DTWTQTDVMTGRTTGTGATYRNNDFFGLVDGLNFVAQYLGKNERDSDFDNTEGNGDGF 197
QY 177 SISYEYEGFCIVGAYGAADRTNLOE-----AQLGNGKKAQWATGLKYDANNIYLAAN 230
Db 198 SATYEYEGFCIVGAYGAADRTNLOE-----AQLGNGKKAQWATGLKYDANNIYLAAN 257
QY 231 YGTRNATPTTNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIGD 290
Db 258 YSETQNTVFADHFV-----ANKAQNFVAQYQDFGLRPSIATYTKSKAKDVEGIGD 312
QY 291 LVNYFEVGYATYYFNKNNSTVVDYIIQIDSDN---KLGVGSDDTVAVGIVYQF 340
Db 313 LVNYFEVGYATYYFNKNNSTVVDYIIQIDSDN---KLGVGSDDTVAVGIVYQF 365

RESULT 12

MMBPP2
outer membrane porin ompC precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 01-Mar-2002
C:Accession: A18885; B25029; E64991
R:Mizuno, T.; Chou, M.Y.; Inouye, M.
J. Biol. Chem. 258, 6932-6940, 1983
A:Title: A comparative study on the genes for three porins of the Escherichia coli ou
A:Reference number: A20867; MUID:83213433; PMID:6304064

A:Accession: A20867
A:Molecule type: DNA
A:Residues: 1-367 <MI2>
A:Cross-references: GB:K00541; GB:M10314; GB:M14188; NID:g147007; PIDN:AAA24243.1; PID:g147007
R:Mizuno, T.; Chou, M.Y.; Inouye, M.
FEBS Lett. 151, 159-164, 1983
A:Title: DNA sequence of the promoter region of the ompC gene and the amino acid sequence of the protein
A:Reference number: A18885; MUID:83132326; PMID:6297988
A:Accession: A18885
A:Molecule type: DNA
A:Residues: 1-40 <MI2>
R:Nogami, T.; Mizuno, T.; Mizushima, S.
J. Bacteriol. 164, 797-801, 1985
A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous recombination
A:Reference number: A91809; MUID:86033642; PMID:2997131
A:Accession: B25029
A:Molecule type: DNA
A:Residues: 32-57 <NOG>
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64991
A:Status: nucleic acid sequence not shown; translation not shown
A:Title: ompC
A:Molecule type: DNA
A:Residues: 1-367 <BLAT>
A:Cross-references: GB:AE000310; GB:U00096; NID:g2367131; PIDN:AAC75275.1; PID:g1788544;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ompC
A:Map position: 47 min
C:Function:
A:Description: one of the E. coli major outer membrane proteins that form passive diffusion pores
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-367/Product: outer membrane porin ompC #status predicted <MAT>

Query Match 58.1%; Score 1051; DB 1; Length 367;
Best Local Similarity 58.9%; Pred. No. 2e-68;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 22 AEVYNKDGKVDLYGKVDGLHYFS-----DNKDVDDQDTYMLRGPKEGTQVTDQLTGYGQ 76

Qy 61 WEYNFQGNNSGADAQTGNKTRLAFAGLKYADVGSFDYGRNYYVVDALGYTDLMPFEGG 120
Db 77 WEYIQGNSAEN---ENNSWTRVAFAGLKFQDVGSFDYGRNYYVVDVTSWTDVLPFEGG 133

Qy 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----A 166
Db 134 DTYGSDFNFMQGRNGFATYRNTDFGLVDGLNFAVQYOGKNGNPSGEGFTSGVTNNGRDA 193

Qy 167 RNSGDDGVGGSISYEGFVGIVGAGAADRTNLQE-ADPLGNKKKAQWATGLKYGANNI 225
Db 194 LRQNGDGVGGSITDYEGFVGIGGAISSSKRTDAQNTAAVIGNDRAETVYTGKLYGANNI 253

Qy 226 YLAANYGETRATPTITNKFTTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDV-E 284
Db 234 YLAAYQTYNATRV-----GSLGWANKAQNFAVQYQDFGLRPSLAYLQSKGNKNGR 308

Qy 285 GIGVDVLYNFEVGAATYFNNKMSYVDYIINQIDSDNKL-----GVGSDTAVAGIVTYQF 340
Db 309 GYDDEDILKYVDVGATYFNNKMSYVDYKINLLD-DNQFTDAGINTDNIVALGLVYQF 367

RESULT 13
S36606
outer membrane porin phoE precursor - Salmonella typhi
C:Species: Salmonella typhi
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999

C:Accession: S36606
R:Torres, A.; Puente, J.L.; Calva, E.
submitted to the EMBL Data Library, August 1993
A:Reference number: S36606
A:Accession: S36606
A:Molecule type: DNA
A:Residues: 1-350 <TOR>
A:Cross-references: EMBL:X74595; NID:g397159; PIDN:CAA52672.1; PID:g397160
A:Experimental source: isolate IMSS-1
A:Note: this is one of the proteins induced when the E. coli cells are grown under phd compounds, and some other negatively charged solutes
C:Genetics:
A:Gene: phoE
C:Complex: homotrimer
C:Function:
A:Description: allows passive diffusion of small hydrophilic molecules with molecular A:Note: induced under phosphate limitation
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-350/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 57.7%; Score 1044; DB 2; Length 350;
Best Local Similarity 57.4%; Pred. No. 6e-68;
Matches 197; Conservative 47; Mismatches 83; Indels 16; Gaps 5;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 21 AEVYNKDGKVDLYGKVKAMHYSDYDSK-----DGDQSYVRFPGKGTQINDLTGYGR 75

Qy 61 WEYNFQGNNSGADAQTGNKTRLAFAGLKYADVGSFDYGRNYYVVDALGYTDLMPFEGG 120
Db 76 WEAEAGNKAESDSQ--QKTRLAFAGLKDIGSFYGRNLGALYDVEAWTDMPEFG 133

Qy 121 D-TAYSDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSIS 179
Db 134 DSAQTDNFMTRKSLATYRNTDFGIYDGLDLDLQYQGNEDRDVKKQNGDGFSTVS 193

Qy 180 YVEYEG--FGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237
Db 194 YDFGSDFAVSAYTSLDRTREQLQRCTGDKAGWATGVKYDANDIYIATFYSETRM 253

Qy 238 TPTITNKFTTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNFEV 297
Db 254 TPV-----SGGFANKTNFQVAVQYQDFGLRPSISLVSKGDIKGVSDLVNVIDV 307

Qy 298 GATYFNNKMSYVDYIINQIDSDNKLGVGSDDTAVAGIVTYQF 340
Db 308 GAIIYFNKMSAFVDYKINQLDSDNTLGINDDDDIIVAILTYQF 350

RESULT 14
MNECNC
outer membrane porin nmpC precursor - Escherichia coli (strain K-12) cryptic lambdaoid
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 01-Mar-2002
A:Accession: A25647; S66594; G64787
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and nmpC outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457; PMID:3017988
A:Contents: mutant strain CS384
A:Accession: A25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
R:Mandl, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A:Title: Holiday junction resolvases encoded by homologous rusA genes in Escherichia A:Reference number: S66579; MUID:96196428; PMID:8648624
A:Accession: S66594
A:Molecule type: DNA
A:Residues: 347-365 <MANH>
A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CAA63325.1; PID:g1051145

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64787
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 'MNIYRATSPFNSSKKGLT', 1-325, 'N', 327-347, 'EGANKSLI' <BLAT>
A;Cross-references: GB:AE000160; GB:U00096; NID:g1786751; PIDN:AACT3654.1; PID:g1786765;
A;Experimental source: strain K-12, substrain MG1655
C;Comment: In wild-type strains of E. coli K-12, the mmpC open reading frame is interrup
ain CS348, the IS5 element has been deleted and mmpC is expressed.
C;Genetics:
A;Gene: mmpC
A;Map position: 12 min
A;Genome: cryptic lambdaoid prophage DLP12
C;Superfamily: outer membrane protein phoE
C;Keywords: membrane protein; porin; trimer
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-365/Product: outer membrane porin mmpC #status predicted <MAT>

Query Match 57.6%; Score 1042; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 8.9e-68;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

QY 1 AEIYKDKGKVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
Db 24 AEIYKDKGKVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 78
QY 61 WEYNFQGNSEGAQAOTGNKTRAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 120
Db 79 WEYEFKGNRAE-SQSSKDKTRAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 137
QY 121 DT-AYSDDFFVGRGVGATYRNSNFFGLVDGLNFAVOYLGNKNE---DTARRSNGDGVGG 176
Db 138 DTWQTQDVFMTQATGVATYRNNDFGLVDGLNFAVOYLGNKNE---DTARRSNGDGVGG 197
QY 177 SISYEYEGFVGIVGAYGAADRTNLOE-----AQPLNGKKAQWATGLKYDANNIYLAAN 230
Db 198 SATYEEGFGIGATYAKSDRTDQVNAAGKLVPEVAFSGKNAEVAAGLKYDANNIYLAAT 257
QY 231 YGETRNATPTTNKFTNTSGFANTQDVLVAQYQDFGLRPSIAYTKSKAKDVEGIGDVD 290
Db 258 YSETQNTVFADHFV-----ANKAQNEFAVAQYQDFGLRPSVAYLQSKGKDLGVWGDQD 312
QY 291 LVNYFEVGCATYFNKNMSTYVDYIINQIDSDN---KLGVGSDDTVAVGIVYQF 340
Db 313 LVKYVDVGATYFKNMSTYVDYKINLLDKNDFKALGVSTDDDIVAVGLVYQF 365

RESULT 15
D64888
outer membrane porin bl377 precursor - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: D64888
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D64888
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-377 <BLAT>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4459.1; PID:g1787641;
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: outer membrane protein phoE
C;Keywords: membrane protein; porin; trimer
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-377/Product: probable outer membrane porin bl377 #status predicted <OPP>

Query Match 57.2%; Score 1035; DB 2; Length 377;
Best Local Similarity 56.9%; Pred. No. 2.9e-67;
Matches 209; Conservative 42; Mismatches 78; Indels 38; Gaps 9;

QY 1 AEIYKDKGKVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
Db 22 AEVYKDKGKVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 76
QY 61 WEYNFQGNSEGAQAOTGNKTRAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 120
Db 77 WEYNIQANNTESSKNQSW--TRLAFAGLKFADYGSFDYGRNYGVYDIEGTMDLPEFGG 134
QY 121 DT-AYSDDFFVGRGVGATYRNSNFFGLVDGLNFAVOYLGNKNE-----RDTAR 167
Db 135 DSYTNADNFWTGRANGVATYRNTDFFGLVNLGNLFAVOYQGNNEGASNGQEGTNNCRD--VR 193
QY 168 RNSGDGVGSGISYBY-EGFGIVGAYGAADRTNLOEAPLNGKKAQWATGLKYDANNIY 226
Db 194 HENGSGGLSTTYDLGNGFSAGAAATSSDRTNDQVNIHTAAGDKADAWTAGLKYDANNIY 253
QY 227 LAANYGETRNATPTTNKFTNTSGFANTQDVLVAQYQDFGLRPSIAYTKSKAKDVE-- 284
Db 254 LATMYSETRNMTPEFD---SDYAVANKTQNPETVAQYQDFGLRPAVSVFLMSKGRDLHAA 310
QY 285 -----GIGDVLVNYFEVGCATYFNKNMSTYVDYIINQIDSDNKL-----GVGSDDTVA 333
Db 311 GGADNPAGVDDKOLVKYADIGATYFNKNMSTYVDYKINLLDEDDSFYAANGISTDDIVA 370
QY 334 VGIVYQF 340
Db 371 LGLVYQF 377

Search completed: December 18, 2002, 16:22:23
Job time : 11.0756 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 : Search time 5.23408 Seconds
(without alignments)
2702.183 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDGKVDLYGKAVGL.....NKLGVGSDTVAVGIYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	OMPFC_ECOLI
2	1110	61.4	351	1	PHOE_ECOLI
3	1077.5	59.6	362	1	OMPFC_ECOLI
4	1072.5	59.3	350	1	PHOE_ENTCL
5	1072	59.3	374	1	OMPFC_SERMA
6	1070	59.1	351	1	PHOE_CITFR
7	1058	58.5	350	1	PHOE_SALTI
8	1058	58.5	350	1	PHOE_SALTY
9	1051	58.1	367	1	OMPFC_ECOLI
10	1050	58.0	351	1	PHOE_KLEPN
11	1046	57.8	365	1	PORI_BPPA2
12	1042	57.6	365	1	NMPC_ECOLI
13	1041	57.5	349	1	PHOE_KLEOX
14	1035	57.2	377	1	OMPFC_ECOLI
15	1031.5	57.0	363	1	OMPFC_SALTI
16	1031.5	57.0	363	1	OMPFC_SALTY
17	1030	56.9	363	1	OMPFC_KLEPN
18	1027	56.8	383	1	OMPFC_SALTI
19	1023.5	56.6	378	1	OMPFC_SALTY
20	999	55.2	397	1	YEDS_ECOLI
21	976.5	54.0	394	1	OMPFC_SALTI
22	902	49.9	342	1	OMPFC_RAHQ
23	894.5	49.4	376	1	OMPFC_SERMA
24	885	48.9	369	1	OMPFC_XENNE
25	573.5	31.7	382	1	PORL_BUCAI
26	236.5	13.1	341	1	OMPFC_VIBCH
27	218	12.1	341	1	OMPFC_PHOPR
28	170.5	9.4	353	1	OP28_HAEIN
29	151.5	8.4	371	1	OP25_HAEIN
30	150.5	8.3	331	1	OMPFC_NEIMB
31	141.5	7.8	360	1	OP26_HAEIN
32	139.5	7.7	359	1	OP21_HAEIN
33	139	7.7	361	1	OP29_HAEIN

ALIGNMENTS

RESULT 1

ID	OMPFC_ECOLI	STANDARD	PRT	362 AA.
AC	P02931;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Outer membrane protein F precursor (Porin ompF) (Outer membrane protein 1A) (Outer membrane protein 1A) (Outer membrane protein B).			
DE	OMPFC OR TOLF OR CMLB OR COA OR CRY OR B0929.			
GN	Escherichia coli.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=K12;			
RC	MEDLINE=83090452; PubMed=6294623;			
RX	Inokuchi K., Mutoh N., Matsuyama S.-I., Mizushima S.;			
RT	"Primary structure of the ompF gene that codes for a major outer membrane protein of Escherichia coli K-12."			
RL	Nucleic Acids Res. 10:6957-6968(1982).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,			
RA	Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,			
RA	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,			
RA	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,			
RA	Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RA	Yano M., Horiuchi T.;			
RT	"A 718-Kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155(1996).			
RN	[4]			
RP	SEQUENCE OF 1-37 FROM N.A.			
RX	MEDLINE=82139379; PubMed=7037455;			
RA	Mutoh N., Inokuchi K., Mizushima S.-I.;			
RT	"Amino acid sequence of the signal peptide of OmpF, a major outer membrane protein of Escherichia coli."			
RL	FEBS Lett. 137:171-174(1982).			
RN	[5]			
RP	SEQUENCE OF 23-362.			
RX	MEDLINE=82256494; PubMed=7049161;			
RA	Chen R., Kramer C., Schmidmayr W., Chen-Schmeisser U., Henning U.;			

P18195 neisseria g
P20149 haemophilus
Q48218 haemophilus
Q9JYV5 neisseria m
P20148 neisseria g
P30688 neisseria m
P55127 neisseria m
Q48219 haemophilus
Q48220 haemophilus
P30687 neisseria m
P30689 neisseria m
Q48253 helicobacte

34 138.5 7.7 348 1 OMB1_NEIGO
35 138.5 7.7 361 1 OP22_HAEIN
36 135.5 7.5 369 1 OP27_HAEIN
37 133 7.4 1829 1 FRPC_NEIMB
38 131.5 7.3 348 1 OMB2_NEIGO
39 131 7.2 331 1 FRPC_NEIMC
40 131 7.2 1829 1 FRPC_NEIMC
41 130 7.2 367 1 OP2A_HAEIN
42 129.5 7.2 363 1 OP2B_HAEIN
43 129 7.1 331 1 OMB1_NEIMB
44 129 7.1 331 1 OMB3_NEIMB
45 127 7.0 1310 1 VAC3_HELPY

RT Primary structure of major outer-membrane protein I (ompF protein,
RL porin) of *Escherichia coli* B/r.";
RN Biochem. J. 203:33-43(1982).
[6]
RX SEQUENCE OF 33-63 FROM N.A.
RA MEDLINE-86033642; PubMed-2997131;
Nogami T., Mizuno T., Mizushima S.;
RT "Construction of a series of ompF-ompC chimeric genes by in vivo
RT homologous recombination in *Escherichia coli* and characterization of
RT the translational products.";
RL J. Bacteriol. 164:797-801(1985).
[7]
RN SEQUENCE OF 23-34 AND 39-47.
RC STRAIN-K12 / EMG2;
RX MEDLINE-97443975; PubMed-9298646;
Link A.J., Robison K., Church G.M.;
RA "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313(1997).
[8]
RN SEQUENCE OF 23-27.
RC STRAIN-K12 / W3110;
RX MEDLINE-98291876; PubMed-9629924;
Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RL Electrophoresis 19:837-844(1998).
[9]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE-92375189; PubMed-1380671;
Cowan S.W., Schirmer T., Rummel G., Steiert M., Ghosh R.,
RA Pauptit R.A., Jansonius J.N., Rosenbusch J.P.;
RT "Crystal structures explain functional properties of two *E. coli*
RT porins.";
RL Nature 358:727-733(1992).
[10]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF MUTANT ASP-141.
RX MEDLINE-95024177; PubMed-7524100;
Jeanteur D., Schirmer T., Fourel D., Simonet V., Rummel G., Widmer C.,
RA Rosenbusch J.P., Pattus F., Pages J.M.;
RT "Structural and functional alterations of a colicin-resistant mutant
RT of OmpF porin from *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:10675-10679(1994).
[11]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE-99060042; PubMed-9843370;
Phale P.S., Philippsen A., Kiefhaber T., Koebnik R., Phale V.P.,
RA Schirmer T., Rosenbusch J.P.;
RT "Stability of trimeric OmpF porin: the contributions of the latching
RT loop L2.";
RL Biochemistry 37:15663-15670(1998).
CC -!- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01655; AAA2444.1; -;
DR EMBL; AE000195; AAC74015.1; -;
DR EMBL; D90730; BAA35675.1; -;
DR EMBL; D90731; BAA35681.1; -;
DR PIR; A03431; MMECF.
DR PIR; A25029; A25029.

DR PDB; 2OMF; 07-DEC-95.
DR PDB; 1OPF; 07-FEB-95.
DR PDB; 1GFM; 07-DEC-96.
DR PDB; 1GFN; 07-DEC-96.
DR PDB; 1GFO; 07-DEC-96.
DR PDB; 1GFP; 07-DEC-96.
DR PDB; 1GFQ; 07-DEC-96.
DR PDB; 1IMP; 07-FEB-95.
DR PDB; 1BT9; 13-JAN-99.
DR SWISS-2DPAGE; P02931; COLI-
DR ECODBASE; B036.0; 6TH EDITION.
DR EcoGene; EG10671; ompF.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW 3D-structure; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 362 OUTER MEMBRANE PROTEIN F.
FT CONFLICT 88 88 Q -> E (IN REF. 5).
FT CONFLICT 139 139 E -> G (IN REF. 5).
FT CONFLICT 284 284 Q -> L (IN REF. 5).
FT STRAND 24 28
FT TURN 29 30
FT STRAND 31 45
FT TURN 50 51
FT STRAND 53 53
FT STRAND 58 59
FT STRAND 62 72
FT STRAND 77 88
FT TURN 95 100
FT STRAND 102 112
FT TURN 113 115
FT STRAND 116 124
FT TURN 126 127
FT HELIX 128 131
FT TURN 132 132
FT TURN 143 144
FT TURN 147 148
FT TURN 150 151
FT STRAND 154 163
FT HELIX 165 169
FT TURN 171 172
FT STRAND 173 180
FT STRAND 183 183
FT TURN 188 190
FT STRAND 192 192
FT TURN 205 206
FT STRAND 207 217
FT HELIX 220 223
FT TURN 224 224
FT STRAND 227 227
FT STRAND 232 244
FT TURN 245 246
FT STRAND 247 257
FT STRAND 261 264
FT TURN 265 268
FT STRAND 269 272
FT STRAND 275 285
FT TURN 288 289
FT STRAND 291 305
FT TURN 306 308
FT STRAND 309 324
FT STRAND 329 338
FT STRAND 351 351
FT STRAND 353 361
SQ SEQUENCE 362 AA; 39333 MW; 3F0974D96DB65464 CRC64;
Query Match 99.8%; Score 1805; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 2.9e-123;

```
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEIYKDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGYG 60
DB 23 AEIYKDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGYG 82
QY 61 WEYNFQGNNEGADQTCNTRLAFAGLKVADVGSFYGNYGVVYDALGVTDMLPFEGG 120
DB 83 WEYNFQGNNEGADQTCNTRLAFAGLKVADVGSFYGNYGVVYDALGVTDMLPFEGG 142
QY 121 DTAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRRSNGDVGGSISY 180
DB 143 DTAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRRSNGDVGGSISY 202
QY 181 EYEGFGIVGAYGAADRTNLOEAQPLNGKKAQWATGLKGDANNIYLAANYGETRNP 240
DB 203 EYEGFGIVGAYGAADRTNLOEAQPLNGKKAQWATGLKGDANNIYLAANYGETRNP 262
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFGAT 300
DB 263 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFGAT 322
QY 301 YFNKNNSTVYDIINOIDSNDKLGVSDDTVAVGIYQF 340
DB 323 YFNKNNSTVYDIINOIDSNDKLGVSDDTVAVGIYQF 362

RESULT 2
PHOE_ECOLI
ID PHOE_ECOLI STANDARD; PRT; 351 AA.
AC P02932;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR OMPE OR B0241.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=83189086; PubMed=6341601;
RA Overbeek N., Bergmans H., van Mansfeld F., Lugtenberg B.;
RT "Complete nucleotide sequence of phoE, the structural gene for the
RT phosphate limitation inducible outer membrane pore protein of
RT Escherichia coli K12."
RL J. Mol. Biol. 163:513-532(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region."
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
```

```
RA Davis R.W.;
RN Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RX SEQUENCE OF 1-22 FROM N.A.
DB MEDLINE=84297232; PubMed=6089111;
RA Deutch A.H., Rushlow K.E., Smith C.J.;
RT "Analysis of the Escherichia coli proBA locus by DNA and protein
RT sequencing."
RL Nucleic Acids Res. 12:6337-6355(1984).
RN [6]
RP MUTAGENESIS OF PHE-351.
RX MEDLINE=91162638; PubMed=1848301;
RA Struyve M., Moons M., Tommassen J.;
RT "Carboxy-terminal phenylalanine is essential for the correct assembly
RT of a bacterial outer membrane protein."
RL J. Mol. Biol. 218:141-148(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (6.0 ANGSTROMS).
RX MEDLINE=91172301; PubMed=1848682;
RA Jap B.K., Walian P.J., Gehring K.;
RT "Structural architecture of an outer membrane channel as determined
RT by electron crystallography."
RL Nature 350:167-170(1991).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=92375189; PubMed=1380671;
RA Cowan S.W., Schlirner T., Rummel G., Steiert M., Ghosh R.,
RA Paupit R.A., Jansonius J.N., Rosenbusch J.P.;
RT "Crystal structures explain functional properties of two E. coli
RT porins."
RL Nature 358:727-733(1992).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=93172954; PubMed=7679770;
RA Struyve M., Visser J., Adriaanse H., Benz R., Tommassen J.;
RT "Topology of PhoE porin: the 'eyelet' region."
RL Mol. Microbiol. 7:131-140(1993).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN E. COLI CELLS
CC ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: V00316; CAA23605.1; -
CC EMBL: AE000132; AAC73345.1; -
CC EMBL: D83536; BAA77910.1; -
CC EMBL: U70214; AAB08661.1; -
CC EMBL: X00786; CAA25362.1; -
CC PIR: A03432; MMECEP.
CC PDB: 1PHO; 31-OCT-93.
CC
CC SWISS-2DPAGE: P02932; COLI.
CC ECO2DBASE: B037.0; 6TH EDITION.
CC EcoGene: EG10729; phoE.
CC InterPro: IPR003229; OMP.
CC InterPro: IPR001702; Porin-gram-ve.
CC Pfam: PF00267; Gram-ve-porins; 1.
CC PRINTS: PR00182; ECOLNEIPORIN.
CC PROSITE: PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal; 3D-structure;
KW Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
```

```
FT MUTAGEN 351 351 F->Y,N,S,V: LESS RESISTANT TO TRYPSIN.
FT STRAND 25 26
FT STRAND 30 35
FT STRAND 37 45
FT STRAND 52 53
FT STRAND 56 65
FT STRAND 71 82
FT TURN 87 88
FT STRAND 93 104
FT TURN 105 107
FT STRAND 108 116
FT TURN 119 119
FT HELIX 120 123
FT TURN 124 126
FT TURN 140 141
FT TURN 143 145
FT STRAND 147 157
FT TURN 159 162
FT TURN 164 165
FT STRAND 166 173
FT STRAND 176 176
FT TURN 182 183
FT STRAND 185 185
FT STRAND 188 197
FT TURN 198 201
FT STRAND 202 212
FT HELIX 215 217
FT TURN 218 219
FT STRAND 227 239
FT TURN 240 241
FT STRAND 242 252
FT STRAND 264 274
FT STRAND 280 292
FT STRAND 294 294
FT TURN 295 297
FT STRAND 298 298
FT STRAND 301 313
FT STRAND 318 327
FT TURN 334 335
FT STRAND 342 350
SQ SEQUENCE 351 AA; 249E2E362C030C9A CRC64;

Query Match 61.4%; Score 1110; DB 1; Length 351;
Best Local Similarity 62.1%; Pred. No. 3.5e-73;
Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKEQINSDLTGXGO 60
DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKEQINSDLTGXGO 76

QY 61 WEYFQGNSEGADQTKNKLAFAGLKYADVGSDYGRNYYGVVYDALGYTDMLEPFGG 120
DB 77 WEAFEPAGNKAESDTAQ--QKTRLAFAGLKYKDLGSDYGRNLGALYDVEAWTDMPEFGG 134

QY 121 D-TAYSDDEFVGRVCGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARNSGDVGGSIS 179
DB 135 DSSAQTDNFMTRKASGLATYRTDFGVIDGLNLTLQYQGNENRDVKKQNGDGFCTSLT 194

QY 180 YEYEG--FGIVGACGAADRTNLQEAQPLGNGKKAQEWATGLKYDANNLYLAANYGETRNA 237
DB 195 YDFGSDFAISGAYYNSDRTNQNLQSRGTGRACAWATGLKYDANNLYLATFYSETRRM 254

QY 238 TPITNKNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKADVEGIDVDLVNYPEV 297
DB 255 TPI-----TGSFANKTNFEAVAQYQDFGLRPSLGYLSKGDIEGIDEDLVNIDV 308

QY 298 GATYFFNKMSTYVDYIINQISDKNLGLVGSDDTVAVGIVTQF 340
DB 309 GATYFFNKMSPVYDKINQLDSDNKLINNDIIVAVGMTYQF 351
```

RESULT 3
OMPD_SALTY

```
ID AC P37592;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane porin protein ompd precursor.
GN OMPD OR NMPC OR STM1572.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 80-362 FROM N.A.
RC STRAIN=SLI303;
RX MEDLINE=95011654; PubMed=7926834;
RA Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA Ichimura S., Noda Y.;
RT "The methyl viologen-resistance-encoding gene smvA of Salmonella
RT typhimurium."
RL Gene 148:173-174(1994).
RN [3]
RP IDENTIFICATION AS OMPD.
RA Singh S.P., Miller S., Williams Y.U., Rudd K.E., Nikaido H.;
RL Unpublished observations (FEB-1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AE008768; AAL20490.1; -.
CC EMBL; D26057; BAA05056.1; -.
CC HSSP; P02931; IGFN.
CC StyGene; SG10249; ompd.
CC InterPro; IPR003229; OMP_2.
CC InterPro; IPR001702; Porin_gram-ve.
CC Pfam; PF00267; Gram-ve_porins; 2.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 362 OUTER MEMBRANE PORIN PROTEIN OMPD.
FT CONFLICT 157 157 N->K (IN REF. 2).
FT CONFLICT 258 258 MISSING (IN REF. 2).
SQ SEQUENCE 362 AA; 39679 MW; 9F68A9A7B948174C CRC64;

Query Match 59.6%; Score 1077.5; DB 1; Length 362;
Best Local Similarity 62.3%; Pred. No. 8e-71;
Matches 218; Conservative 31; Mismatches 82; Indels 19; Gaps 7;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKEQINSDLTGXGO 60
DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKEQINSDLTGXGO 76

QY 61 WEYFQGN--NSEGADQTKNKLAFAGLKYADVGSDYGRNYYGVVYDALGYTDMLEPFG 118
DB 77 WEYFQGNRTESQGADK---DKTRLAFAGLKFADYGSFDYGRNYYGVVYDALGYTDMLEPFG 133
```

QY 119 GGDY-AYSDFFYGRVGGVATYRNSFFGLVDGLNFAVOYLGNKRDRTARRSNGDVGGS 177
DB 134 GDTWTQTDVFTVGRGTGATYRNTDFGLVEGLNFAAQYQGNKRDGAYESNGDGFGLS 193
QY 178 ISYEEFGFIVGAYGAARDTNLQEAQPLGN----GKKAQWATGLKYDANNIYLAANYGE 233
DB 194 ATYEEFGFVGAAYAKSDRTN-NOVKAASNLNAAGNAEWAAGLADANNIYLAANYGE 252
QY 234 TRNATITNKFNTSFGANKTQDVLVLAQYQDFGLRPSIATYKSAKOVGEGDVLVN 293
DB 253 TLMNTTFGDAADAPAIKNTQNEFAVAQYQDFGLRPSIATYKSAKOVGEGDVLVN 312
QY 294 YFVGATYFKNMSTYVDYIINOISDN---KLGVGSDDTVAAGVIYQF 340
DB 313 YIDVGATYFKNMSTYVDYIINOISDN---KLGVGSDDTVAAGVIYQF 362
RESULT 4
PHOE_ENTCL STANDARD; PRT; 350 AA.
AC 047490;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Enterobacter cloacae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tomassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475(1987).
CC -!- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28296; AAA24809.1; -
DR HSSP: P02932; 1PHO.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gran-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 350 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 350 AA; 38511 MW; 101665FEECC3AFA4 CRC64;
Query Match 59.3%; Score 1072.5; DB 1; Length 350;
Best Local Similarity 60.3%; Pred. No. 1.8e-70;
Matches 207; Conservative 40; Mismatches 79; Indels 17; Gaps 5;
QY 1 AEYNGKGNVDLYGKAVGLHYFKSCNGENSYGGNGDMYIARLFGKGETQINSDLTGCG 60

DB 22 AEYVNGKGNKLDYGVKAMHYISDDTK-----DCDQTVVRFCKGETQINOLTYGR 76
QY 61 WEYVNGKGNSEGAADATGKNTLAFAGLKYADYVGSFEDYGRNYGVYDALGYTDLMPFEG 120
DB 77 WEAEFAGNAESDSSQ---KTRLAFAAGLKLKFGSLDYGRNLGALYDEAWTDMFPEFG 133
QY 121 D-TAYSDFFYGRVGGVATYRNSFFGLVDGLNFAVOYLGNKRDRTARRSNGDVGGSIS 179
DB 134 DSSAQIDNFTWKRAASGLATYRNTDFGLDGLMTLQYQGNKRNDRDAKONGDGFGLSLT 193
QY 180 YEYEG--FGIVGAYGAARDTNLQEAQPLGNKGAQWATGLKYDANNIYLAANYGETRNA 237
DB 194 YDFGGTDFAVSGAYTNSDRINAQNLARAQGAQWATGLKYDANDIYLAANYSETRNM 253
QY 238 TPITNKTNTSFGANKTQDVLVLAQYQDFGLRPSIATYKSAKOVGEGDVLVNYFEY 297
DB 254 TPI-----SGGFANKAQNPEVVAQYQDFGLRPSIATYKSAKOVGEGDVLVNYFEY 307
QY 298 GATYFKNMSTYVDYIINOISDNKLGVGSDDTVAAGVIYQF 340
DB 308 GATYFKNMSTYVDYIINOISDNKLGVGSDDTVAAGVIYQF 350
RESULT 5
OMP_FSERMA STANDARD; PRT; 374 AA.
AC 033980;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane protein F precursor (Porin ompF).
GN OMPF.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCC-51;
RX MEDLINE=97419518; PubMed=9274033;
RA Huttsul J.A.M., Worobec E.A.;
RT "Molecular characterization of the Serratia marcescens OmpF porin,
RT and analysis of S. marcescens OmpF and OmpC osmoregulation.";
RT Microbiology 143:2797-2806(1997).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U01967; AAB69103.1; ALT_INIT.
DR HSSP: Q48473; IOSM.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gran-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 374 OUTER MEMBRANE PROTEIN F.
SQ SEQUENCE 374 AA; 41184 MW; BF821D385E441049 CRC64;
Query Match 59.3%; Score 1072; DB 1; Length 374;
Best Local Similarity 60.1%; Pred. No. 2.1e-70;
Matches 215; Conservative 38; Mismatches 81; Indels 24; Gaps 10;

```
Qy 1 AEIYNKDNKVDLYGKAVGLHYFSGKNGENSYGGNDMTYARLGFPGKETQINSDLTGYGQ 60
Db 23 AEIYNKDNKLDLYGKVDGLHYFSGKDK-----GNDQDTYVRFPGKETQITDQLTGYGQ 77
Qy 61 WEYNFQGNSSGADQATGNKTRLAFAGLYADVGSDYGRNYYVYDYGALGYDMLPERGG 120
Db 78 WEYNVQSHASQGTG-GTKRLGAGLKFADYGSFDYGRNYYVLYDVEGWTDMPLPERGG 136
Qy 121 DT-AISDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERD--TARRSNGDGVGGS 177
Db 137 DTYTSDNFMGTGRNGVATYRNNFFGLVDGLNFAVQYLGKNERD--TARRSNGDGVGGS 196
Qy 178 ISYEF-EGFGLVGVAGAADRNQLQAQPLGNGKKAQWATGLKYDANNIYLAANYGETRN 236
Db 197 STYDIGEVSFGAAYASSNRDQKLRSENERGDKADATVGAKYDANNVYLAANYAETRN 256
Qy 237 ATP1-TNKTNT-----SGFANKTQVLLVAQYQDFGLRPSIATYKSKAK--DVSGI 286
Db 257 MTPFGGGNFTTCAATENCGGFASKTQNFVTAQYQDFGLRPSIATYKSKAK--DVSGI 316
Qy 287 G-DVDLVNVEFGVATYRNNKSTVVDYIINQIDSDN---KLGVGSDDTVAVGIVYQF 340
Db 317 GSDQDLVKYVSGVITYYFNKKNSTVVDYIINQIDSDN---KLGVGSDDTVAVGIVYQF 374

RESULT 6
PHOE_CITFR
ID PHOE_CITFR STANDARD; PRT; 351 AA.
AC Q01605;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Sierlings G., Ockhuijsen C., Hofstra H., Tommassen J.;
RT "Characterization of the Citrobacter freundii phoE gene and
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X68021; CAA48162.1; -
CC PIR; S25520; S25520.
CC DR HSP; P02932; 1PHO.
CC DR InterPro; IPR003229; OMP_2.
CC DR Pfam; PF00267; Gram-ve_porins; 1.
CC DR PRINTS; PR00182; ECOLNEIPORIN.
CC DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC KW Outer membrane; Transmembrane; Porin; Signal.
CC FT SIGNAL 1 21
CC CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
```

```
SO SEQUENCE 351 AA; 38807 MW; 4BE9C144070381B5A CRC64;
Query Match 59.1%; Score 1070; DB 1; Length 351;
Best Local Similarity 59.2%; Pred. No. 2.7e-70;
Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;

Qy 1 AEIYNKDNKVDLYGKAVGLHYFSGKNGENSYGGNDMTYARLGFPGKETQINSDLTGYGQ 60
Db 22 AEIYNKDNKLDLYGKVKAMHYMTDYDSK-----DGDSYIRLGFPGKETQINDELTYGGR 76
Qy 61 WEYNFQGNSSGADQATGNKTRLAFAGLYADVGSDYGRNYYVYDYGALGYDMLPERGG 120
Db 77 WEAEFAGNAKESDSNQ--QKTRLAFAGSKLNLGSDYGRNLGALYDVEATDMFPERGG 134
Qy 121 D-TAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERD--TARRSNGDGVGGSIS 179
Db 135 DSSAQTDNFMKTRASGLATYRNTDFGVVDGLDLTLQYQGNQDRDVKKQNGDGFSTV 194
Qy 180 YEYEG--FGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237
Db 195 YDFGSDFAVSGAVTNSDRTNQNLQTRGTGDKAEAWATGLKYDANDIYIATFYSETRNM 254
Qy 238 TPITNKFTNSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVLDVNYFEV 297
Db 255 TPI-----SGFANKTQNFVAVYQYQDFGLRPSIATYKSKAKDVEGIDVLDVNYIDV 308
Qy 298 GATYFNKKNSTVVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
Db 309 GATYFNKKNSTVVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 351

RESULT 7
PHOE_SALTI
ID PHOE_SALTI STANDARD; PRT; 350 AA.
AC Q56119;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR STY0365.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IMSS-1;
RA Torres A., Puente J.L., Calva E.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X74595; CAA52672.1; -;
EMBL; AL627266; CAD08790.1; -;
HSP; P02932; IPHO.
InterPro; IPR003229; OMP_2.
InterPro; IPR001702; Porin-gram-ve.
Pfam; PF00267; Gram-ve porins; 1.
PRINTS; PR00182; ECOLNEIPORIN.
PROSITE; PS00576; GRAM_NEG_PORIN; 1.
Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY
FT CHAIN 21 350 OUTER MEMBRANE PORE PROTEIN E.
FT INTERPRO IPR003229; OMP_2.
FT CONFLICT 63 63 E > K (IN REF. 1).
FT CONFLICT 80 80 F > G (IN REF. 1).
FT CONFLICT 229 229 A > G (IN REF. 1).
SEQUENCE 350 AA; 38744 MW; 890F5F8C5EF0C77 CRC64;

Query Match 58.5%; Score 1058; DB 1; Length 350;
Best Local Similarity 58.0%; Pred. No. 2e-69;
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEYKDGKNDVLYGKAVGLHYFSKNGENSGYGGDMTYARLGFKGTQINSDLTGYGQ 60
DB 21 AEYKNGKNDVLYGKAVGLHYFSKNGENSGYGGDMTYARLGFKGTQINSDLTGYGQ 75
QY 61 WEYFNQGNSEGAQDTGNKTRAFAGLVKADYGVSGFYGRNYGVVYDALGYTMDLPFEGG 120
DB 76 WEAFAGNKAESDSQ--QKTRAFAGLVKADYGVSGFYGRNYGVVYDALGYTMDLPFEGG 133
QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRARRSGDVGGSIS 179
DB 134 DSSAQDTNFTKRSGLATYRNTDFGIVDGLDLTLQYQKNERDTRARRSGDVGGSIS 193
QY 180 YEYEG--FGIVGAGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237
DB 194 YDFGGSDFAVSGAYTLDSTRQNLQRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253
QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADVEIGDVLVNYFEV 297
DB 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSLGLVLSKGDIEGVGSDLVNVIDY 307
QY 298 GATYFNNKNSVYDIYINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 308 GATYFNNKNSAFVDYKINQIDSDNTLGINDDDIVAIGLTYQF 350

RESULT 8
PHOE_SALTY STANDARD; PRT; 350 AA.
AC P30705;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR STM0320
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=93083994; PubMed=1280609;
RA Spierings G., Elders R., van Lith B., Hofstra H., Tommassen J.;
RT "Characterization of the Salmonella typhimurium phoE gene and development of Salmonella-specific DNA probes.";

Gene 122:45-52(1992).

RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RL Nature 413:852-856(2001).
CC 1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE, PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED SOLUTES.
CC 1- SUBUNIT: HOMOTRIMER.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC 1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X68023; CAA48164.1; -;
EMBL; AE008709; AAL19276.1; -;
PIR; S25525; S25525.
HSP; P02932; IPHO.
DR StyGene; SG10291; phoE.
DR InterPro; IPR003229; OMP_2.
DR Pfam; PF001702; Porin-gram-ve.
DR PRINTS; PR00267; Gram-ve porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 350 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 350 AA; 38762 MW; 9D1E3355AF59877 CRC64;

Query Match 58.5%; Score 1058; DB 1; Length 350;
Best Local Similarity 58.0%; Pred. No. 2e-69;
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEYKDGKNDVLYGKAVGLHYFSKNGENSGYGGDMTYARLGFKGTQINSDLTGYGQ 60
DB 21 AEYKNGKNDVLYGKAVGLHYFSKNGENSGYGGDMTYARLGFKGTQINSDLTGYGQ 75
QY 61 WEYFNQGNSEGAQDTGNKTRAFAGLVKADYGVSGFYGRNYGVVYDALGYTMDLPFEGG 120
DB 76 WEAFAGNKAESDSQ--QKTRAFAGLVKADYGVSGFYGRNYGVVYDALGYTMDLPFEGG 133
QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRARRSGDVGGSIS 179
DB 134 DSSAQDTNFTKRSGLATYRNTDFGIVDGLDLTLQYQKNERDTRARRSGDVGGSIS 193
QY 180 YEYEG--FGIVGAGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237
DB 194 YDFGGSDFAVSGAYTLDSTRQNLQRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253
QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADVEIGDVLVNYFEV 297
DB 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSLGLVLSKGDIEGVGSDLVNVIDY 307
QY 298 GATYFNNKNSVYDIYINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 308 GATYFNNKNSAFVDYKINQIDSDNTLGINDDDIVAIGLTYQF 350

```

RESULT 9
OMPC_ECOLI
ID OMPC_ECOLI STANDARD; PRT; 367 AA.
AC P06996;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein C precursor (Porin ompC) (Outer membrane
DE protein 1B).
GN OMPC OR MEOA OR PAR OR B2215.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83213433; PubMed=6304064;
RA Mizuno T., Chou M.-Y., Inouye M.;
RT "A comparative study on the genes for three porins of the Escherichia
RT coli outer membrane. DNA sequence of the osmoregulated ompC gene.";
RL J. Biol. Chem. 258:6932-6940(1983).
[2]
RN SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RX STRAIN-K12;
RC MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
[4]
RN SEQUENCE OF 218-367 FROM N.A.
RC STRAIN-K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 1-22 FROM N.A., AND SEQUENCE OF 22-40.
RX MEDLINE=83132326; PubMed=6297988;
RA Mizuno T., Chou M.-Y., Inouye M.;
RT "DNA sequence of the promoter region of the ompC gene and the amino
RT acid sequence of the signal peptide of pro-OmpC protein of
RT Escherichia coli.";
RL FEBS Lett. 151:159-164(1983).
[6]
RN SEQUENCE OF 32-57 FROM N.A.
RX MEDLINE=86033642; PubMed=2997131;
RA Nogami T., Mizuno T., Mizushima S.;
RT "Construction of a series of ompF-ompC chimeric genes by in vivo
RT homologous recombination in Escherichia coli and characterization of
RT the translational products.";
RL J. Bacteriol. 164:797-801(1985).
[7]
RN SEQUENCE OF 22-30.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded

```

```

RT In the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
[8]
RN SEQUENCE OF 22-26.
RX STRAIN-K12 / W3110;
RC MEDLINE=98291876; PubMed=9629924;
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RL Electrophoresis 19:837-844(1998).
CC -!- FUNCTION: FORM PASSIVE DIFFUSION PORES WHICH ALLOW SMALL MOLECULAR
CC WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K00541; AAA24243.1; -.
DR EMBL; AE000310; AAC75275.1; -.
DR EMBL; D90850; BAA15998.1; -.
DR EMBL; U00008; AAL16412.1; -.
DR PIR; A20867; MWEPC.
DR PIR; B25029; B25029.
DR PIR; A18885; A18885.
DR HSP; O52503; IIIV.
DR ECO2DBASE; A035_5; 6TH EDITION.
DR EcoGene; EG10670; ompC.
DR InterPro; IPR003229; OMP_2.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 367
FT CDS 22 367
SQ SEQUENCE 367 AA; 40368 MW; 6A49370CC8A1A225 CRC64;
-----
Query Match 58.1%; Score 1051; DB 1; Length 367;
Best Local Similarity 58.9%; Pred. No. 6.6e-69;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
-----
QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNGDMTYARLFGKGTQINSDLTGYGQ 60
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 AEVYNKDGKLDLYGKVDGLHYFS-----DNKDVGDQDTYMRGLGFGETQVTDLTGYGQ 76
QY 61 WEYFNQGNNSGADATQGNKTLAFAGLYADYAGSFYDGRNYSVYDYLALGYDMLPEFG 120
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 WEYIQGNSAEN---ENNSWTRVAFAGLKFQDVGSFYSGRNYSVYDVTWTDVLPFGG 133
QY 121 DPAYSDDFEVGRVGVGATYRNPFGLVDGLNFAVQYLCKNERDT-----A 166
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 DTYGSDNFMNQQRNGNFATYRNTDFGLVDGLNFAVQYQKNGNPGSGEGTSGVTNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFIVGAYCAADRTNLQE-AOPLGNGKKAQWATLKKYDANNI 225
DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 LRQNGDVGGSITYDEYEGFICGAISSSKRTDAQNTAAIYICNGDRAEYTGGLKDYANNI 253
QY 226 YLAANYGETRNATPTTNKFTNTSGFANKTDQVLLVAQYQDFGLRPSIAYTKSKARDV-E 284
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 YLAAQYTOTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSLAYLQSKGNLGR 308
QY 285 GIGVDVLYNVEVGATYTFYFNKNMSTYVDYIINQIDSDNKL-----CVGSDDTVAVGIVQF 340
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 GYDDEDILKYVDVGTATYTFYFNKNMSTYVDYKINLLD-DNQFTRDAGINTDNIVALGLVYQF 367

```

```
RESULT 10
PHOE_KLEPN STANDARD; PRT; 351 AA.
AC P30704;
AT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475 (1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M28295; AAA25121.1; -
CC HSSP; P02932; IPHO.
CC InterPro; IPR003229; OMP_2.
CC Pfam; PF00267; Gram-ve_porins; 1..
CC PRINTS; PR00182; ECOLNEIPORIN.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 21 BY SIMILARITY.
CC CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
CC SEQUENCE 351 AA; 38894 MW; 86A5286C12502EC7 CRC64;
Query Match 58.0%; Score 1050; DB 1; Length 351;
Best Local Similarity 58.6%; Pred. No. 7.4e-69;
Matches 201; Conservative 43; Mismatches 83; Indels 16; Gaps 5;
QY 1 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEQTQINSDLTGYGQ 60
DB 22 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEQTQINSDLTGYGQ 76
QY 61 WEYFQGNSEGAADQTKNTRLAFAGLYADVGSFDYGRNYSWYDALGYTDLMLPEFGG 120
DB 77 WESEFSGNKTESDSSQ--QKTRLAFAGVLYKNGYSGFDYGRNGLALYDVAWTDMLPEFGG 134
QY 121 D-TAYSDDFVGRVGVVATYRNSNFFGLVDGLNFAVOYLGNKNERDPAKRSNGDGVGSGIS 179
DB 135 DSSAQTDNFWTKRASGLATYRNTDFGLVDGLTLQYQCKNGREAKKQNGDGVCTSL 194
QY 180 YEYEG--FGVYAGYAADRTNLQEAQPLNGKKAQWATGLKDYANNIYAAYGETRNA 237
DB 195 YDFGGTDFVSAAYTSDDRTNDQNLRLAQAQSKAEAWATGLKYDANNIYATYSETRM 254
QY 238 TPITNFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKADKVEGIDVDLVNFEV 297
```

```
DB 255 TPI-----SGGFANKAQNEEVAQYQDFGLRPSLGYVLSKGDIEGSEDLVNYIDV 308
QY 298 GATYYFNKNKNTYVDYIINQIDSNKLGVSDDTVAVGVIYQF 340
DB 309 GLTYYFNKNMNAFVDYKINQKLSKNGKLGINDDDIVALCMYQF 351

RESULT 11
PORI_BPPA2 STANDARD; PRT; 365 AA.
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane porin protein LC precursor.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC unclassified Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and mmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage.";
RL J. Biol. Chem. 261:12723-12732(1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (by similarity).
CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYCOGENIC STATE IN E. COLI. THE EXPRESSION OF OMPc AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J02580; AAA32301.1; -
CC PIR; D25647; MMBPE2.
CC HSSP; O52503; LIIV.
CC InterPro; IPR003229; OMP_2.
CC InterPro; IPR001702; Porin-gram-ve.
CC Pfam; PF00267; Gram-ve_porins; 1.
CC PRINTS; PR00182; ECOLNEIPORIN.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 23
CC CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
CC CONFLICT 99 99 H -> R (IN REF. 1; AA SEQUENCE).
CC SEQUENCE 365 AA; 40290 MW; 0FBC0531F99C0205 CRC64;
Query Match 57.8%; Score 1046; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 1.5e-68;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;
QY 1 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEQTQINSDLTGYGQ 60
DB 24 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEQTQINSDLTGYGQ 78
QY 61 WEYFQGNSEGAADQTKNTRLAFAGLYADVGSFDYGRNYSWYDALGYTDLMLPEFGG 120
DB 111 WEYFQGNSEGAADQTKNTRLAFAGLYADVGSFDYGRNYSWYDALGYTDLMLPEFGG 120
```


P77747; p76854;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DR Outer membrane protein N precursor (Porin ompN).
DN OMPN OR B137.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
Makiino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nasahimoto H., Nishio Y., Oshima T., Saito N.,
Sampel G., Seki Y., Sivasmaram S., Tagami Y., Takeda J.,
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE FROM N.A.. SEQUENCE OF 22-33, AND CHARACTERIZATION.
RX MEDLINE=98317278; PubMed=9642192;
RA Prilipov A., Phale P.S., Koebnick R., Widmer C., Rosenbusch J.P.;
RT Identification and characterization of two quiescent porin genes,
nmpC and ompN, in Escherichia coli BE";
RL J. Bacteriol. 180:3388-3392(1998).
CC -J- FUNCTION: NON-SPECIFIC PORIN.
CC -I- SUBUNIT: HOMOTRIMER.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -I- SIMILARITY: BELONGS TO THE OMP/C/PHEO FAMILY OF PORINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AE000234; AAC74459.1; .
EMBL; D90775; BAAL14981.1; .
EMBL; D90776; BAAL14986.1; .
HSPG; P02931; IGFN
ECOGene; EG13375; ompN.
InterPro; IPR003229; OMP_2.
InterPro; IPR001702; Porin_gram-v.
Pfam; PF00267; Gram-v_poring; 1.
PRINTS; PR00182; ECOLNEIPORIN.
KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 377 OUTER MEMBRANE PROTEIN N.
SQ SEQUENCE 377 AA; 41220 MW; 442D02CB9BC10F95 CRC64;

Query Match 57.2%; Score 1035; DB 1; Length 377;
Best Local Similarity 56.9%; Pred.No. 9.7e-68;
Matches 209; Conservative 42; Mismatches 78; Indels 38; Gaps 9;

QY 1 ATETYNKGDKGVLDYGKAVALGHLYFSKGNCSGGNGDMTYARLGFKGTQNSDLTGCG 60

Db 22 AEVYKNDGKLDLYCKVDGLHYFSDNSAK-----DGDQSYARLGFKGETQINDQITGYGQ 76
Qy 61 WEYNFQGNNEGADQATGNKTRAFAGLKYADVGSFSDYGRNYSVYDGLGYTMDLPEFGG 120
Db 77 WEYNIQANNESSKQSW--TRLAFAGLKFDYGSFSDYGRNYSVYDGLGYTMDLPEFGG 134
Qy 121 DT-AVSDDFVGRVGVATYRNSNFFGLVDGLNFAVQVYLKNE-----RDTAR 167
Db 135 DSYTNADNFMTGRANGVATYRTDFGLVGLNFAVQVYQGNNEGASNGQEGTNGRD-VR 193
Qy 168 RNSGDSVGSISYEY-EGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIY 226
Db 194 HENGDSGLSTYDILGMGFSAGAAVTSDDRTNDQVNHHTAAGDKADAWTGLKYDANNIY 253
Qy 227 LAANGTETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAVTKSKAKDVE-- 284
Db 254 LATMYSEIRNMTFFGD---SDYAVANKQNFETVTAQYQDFGLRPAVSELSKGRDLHNA 310
Qy 285 -----GIGVDLVNVEFGVATYRNSNFFGLVDGLNFAVQVYLKNE-----RDTAR 333
Db 311 GGADNPAGVDDKDLVKYADIGATYRNSNFFGLVDGLNFAVQVYLKNE-----RDTAR 370
Qy 334 VGIVYQF 340
Db 371 LGLVYQF 377

RESULT 15
OMPF_SALTI
ID OMPF_SALTI STANDARD; PRT; 363 AA.
AC Q56113;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein F precursor (Porin ompF) (Outer membrane protein S3).
GN OMPF OR OMP53 OR STV1002.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IMSS-1;
RA Fernandez-Mora M., Calva E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Quail M., Rutherford K., Leather S., Moule S., O'Gaora P., Parry C.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2 (BY
CC SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPF/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89757; CAA61905.1; -.
DR EMBL; AL627268; CAD05399.1; -.
DR HSSP; P02931; IGFN.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW Complete proteome.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 363 OUTER MEMBRANE PROTEIN F.
FT CONFLICT 307 307 D -> G (IN REF. 1).
SQ SEQUENCE 363 AA; 40106 MW; F5059B37EA516859 CRC64;

Query Match 57.0%; Score 1031.5; DB 1; Length 363;
Best Local Similarity 57.6%; Pred. No. 1.7e-67;
Matches 200; Conservative 54; Mismatches 80; Indels 13; Gaps 8;

Qy 1 AEIYNKDGKVDLYKAVGLHYFSGKNGNSYGGNGDMTYARLGFKGETQINDLTGYGQ 60
Db 23 AEIYNKDGKLDLYKAVGRHVWTT-TGDSK---NADQTYAQIGFGETQINTDLTGFGQ 78
Qy 61 WEYNFQGNNEGADQATGNKTRAFAGLKYADVGSFSDYGRNYSVYDGLGYTMDLPEFGG 120
Db 79 WEYRTKADRAEG-EQONSNLVRLAFAGLKYAEVGSIDYGRNYSVYDGLGYTMDLPEFGG 137
Qy 121 DT---AYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQVYLKNE--NGDGVGG 176
Db 138 ETWGGAYTDNYMTSRAGLLTYRNSDFFGLVDGLSFGIOYQCKNQDNHISINQSDGVGY 197
Qy 177 SISYEYEGFIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRN 236
Db 198 TMAYEFDGFGVTAAYSNSKRTNDQDRD--GNGDRAESWAVGAKYDANNVYLAAYVETRN 256
Qy 237 ATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAVTKSKAKDVEGI-CDVLDVNYF 295
Db 257 MSIVENTVTDTVEMANKTONLEVVAQYQDFGLRPAISVQSKQLNGADGSADLAKYI 316
Qy 296 EVGATYFNKNNMSTYVDYIINQIDSDNKLIG--VGSDDTVAVGIVYQF 340
Db 317 QAGATYFNKNNMSTYVDYIINQIDSDNKLIG--VGSDDTVAVGIVYQF 363

Search completed: December 18, 2002, 16:18:23
Job time : 6.23408 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 20.0203 Seconds
(without alignments)
3509.535 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDNKVDLYGKAVGL.....NKLGVGSDDTVAVGIVYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1789	98.9	362	16 Q8XDF1	Q8xdf1 escherichia
2	1390.5	76.9	351	2 Q93K98	Q93k98 enterobacte
3	1113	61.5	351	16 Q8X7N5	Q8x7n5 escherichia
4	1082	59.8	371	16 Q8ZGS6	Q8zgs6 yersinia pe
5	1061	58.7	363	2 Q93K99	Q93k99 enterobacte
6	1044.5	57.7	374	2 Q87754	Q87754 klebsiella
7	1036	57.3	377	2 Q85030	Q85030 escherichia
8	1035.5	57.2	372	2 Q93K39	Q93k39 klebsiella
9	1033	57.1	367	16 Q8XE41	Q8xe41 escherichia
10	1029.5	56.9	364	2 Q9K597	Q9k597 escherichia
11	1029.5	56.9	378	2 Q9K3E6	Q9k3e6 salmonella
12	1028.5	56.9	366	2 Q9RH85	Q9rh85 escherichia
13	1028.5	56.5	333	2 Q9RAW3	Q9raw3 klebsiella
14	1022.5	56.5	333	2 Q9K3E7	Q9k3e7 salmonella
15	1022.5	56.5	363	2 Q9K3E7	Q9k3e7 salmonella
16	1021.5	56.5	333	2 Q9RR59	Q9rr59 klebsiella

17	1019.5	56.4	333	2 Q9S613	Q9s613 klebsiella
18	1019.5	56.4	333	2 Q8VDE7	Q8vde7 klebsiella
19	1019.5	56.4	333	2 Q8VPA5	Q8vpa5 klebsiella
20	1017.5	56.2	360	2 Q9F889	Q9f889 salmonella
21	1017	56.2	377	16 Q8ZPL4	Q8zpl4 salmonella
22	1007	55.7	375	2 Q9AGC9	Q9agc9 enterobacte
23	1005	55.6	375	2 Q9ALY0	Q9aly0 enterobacte
24	1000	55.3	375	2 Q9AGD1	Q9agd1 enterobacte
25	999	55.2	375	2 Q9AGC8	Q9agc8 enterobacte
26	998.5	55.2	366	16 Q8XAS0	Q8xas0 escherichia
27	998	55.2	375	2 Q93T24	Q93t24 enterobacte
28	995.5	55.0	376	2 Q9AGD0	Q9agd0 enterobacte
29	992	54.8	360	16 Q8ZG94	Q8z94 yersinia pe
30	984.5	54.4	398	16 Q8ZNS7	Q8zns7 salmonella
31	960	53.1	374	16 Q8ZGR1	Q8zgr1 yersinia pe
32	946	52.3	366	2 Q93SG6	Q93sg6 clostridium
33	938	51.9	342	2 Q8RLH4	Q8rlh4 salmonella
34	934.5	51.7	359	2 Q87753	Q87753 klebsiella
35	933	51.6	342	2 Q8RLH5	Q8rlh5 salmonella
36	933	51.6	342	2 Q8RLH3	Q8rlh3 salmonella
37	919	50.8	372	16 Q8ZPH7	Q8zph7 salmonella
38	815	45.1	315	2 Q9EXH8	Q9exh8 klebsiella
39	770	42.6	315	2 Q9ZC71	Q9zc71 yersinia pe
40	649.5	35.9	255	16 Q8X900	Q8x900 escherichia
41	539.5	29.8	224	16 Q8XB87	Q8xb87 escherichia
42	428.5	23.7	130	2 Q5Z641	Q5z641 rannella aq
43	423	23.4	191	16 Q8XB86	Q8xb86 escherichia
44	372.5	20.6	123	16 Q8X9Q2	Q8x9q2 escherichia
45	310	17.1	111	2 P94857	P94857 klebsiella

ALIGNMENTS

RESULT 1

Q8XDF1	PRELIMINARY;	PRT;	362 AA.
ID Q8XDF1			
AC Q8XDF1			
DT 01-MAR-2002 (TremBLrel. 20, Created)			
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)			
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)			
DE Outer membrane protein la (Ia,b,F).			
GN OMPF OR 21276 OR ECS1012.			
OS Escherichia coli O157:H7.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia.			
OX NCBI_TaxID=83334;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX MEDLINE=21074935; PubMed=11206551;			
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,			
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA Welch R.A., Blattner F.R.;			
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL Nature 409:529-533(2001).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / RMD 0509952;			
RX MEDLINE=21156331; PubMed=11258796;			
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL DNA Res. 8:111-22(2001).			
DR EMBL; AF005283; AAC55414.1; -			
DR EMBL; AP002553; BAB34435.1; -			
DR InterPro; IPR003229; OMP_2.			

DR	InterPro: IPR001702; Porin_gram-ve.
DR	Pfam: PF00267; Gram-ve_porins; 1.
DR	PRINTS: PR00182; ECOLNEIPORIN.
DR	ProDom: PD000808; OMP_2; 1.
DR	PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW	Complete proteome.
SQ	SEQUENCE 362 AA; 39361 MW; 9E6ACF4B9DAA8214 CRC64;
Query Match	
Best Local Similarity 98.9%; Score 1789; DB 16; Length 362;	
Matches 338; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGGNDMTYARLGKGETQINSDLITGYGQ 60
DB	23 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGGNDMTYARLGKGETQINSDLITGYGQ 82
QY	61 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDFGYGRNYYGVVYDALGYTDMLPFEGG 120
DB	83 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDFGYGRNYYGVVYDALGYTDMLPFEGG 142
QY	121 DPAYSDDFFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 180
DB	143 DPAYSDDFFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 202
QY	181 EYEGGIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 240
DB	203 EYEGGIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 262
QY	241 TNKFTNSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 300
DB	263 TNKFTNSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 322
QY	301 YFENKMTSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB	323 YFENKMTSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362
RESULT 2	
ID	Q93K98 PRELIMINARY; PRT; 351 AA.
AC	Q93K98;
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;
DT	01-DEC-2001 (Tremblrel. 19, Created)
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	OMP porin precursor.
CN	OMP.
OS	Enterobacter cloacae.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Enterobacter.
OX	NCBI_TaxID=550;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC13047;
RA	Ballesteros F., Domenech-Sanchez A., Martinez-Martinez L., Pascual A.,
RA	Conejo M.C., Benedi V.J.;
RA	"Genetic characterization and role in ss-lactam resistance of the OmpC
RT	and ompF enterobacter cloacae porins.";
RL	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AJ316540; CAC48383.1; -.
DR	InterPro: IPR003229; OMP_2.
DR	InterPro: IPR001702; Porin_gram-ve.
DR	Pfam: PF00267; Gram-ve_porins; 1.
DR	ProDom: PD000808; OMP_2; 1.
DR	PROSITE: PS00576; GRAM_NEG_PORIN; UNKNOWN_1.
KW	Porin; Signal.
FT	SIGNAL
SQ	SEQUENCE 351 AA; 38465 MW; 045617CB302F8968 CRC64;
Query Match	
Best Local Similarity 76.9%; Score 1390.5; DB 2; Length 351;	
Matches 263; Conservative 23; Mismatches 43; Indels 11; Gaps 2;	
QY	1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGGNDMTYARLGKGETQINSDLITGYGQ 60

DB	23 AEIYNKDGKVDLYGKAVGLHYFSND-----GNDGDKTYARLGKGETKINDQLTGIGQ 77
QY	61 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDFGYGRNYYGVVYDALGYTDMLPFEGG 120
DB	78 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDFGYGRNYYGVVYDALGYTDMLPFEGG 137
QY	121 DPAYSDDFFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 180
DB	138 DPAYSDDFFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 197
QY	181 EYEGGIVGAYGAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 240
DB	198 DPAYSDDFFVGRVGVVATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 257
QY	241 TNKFTNSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 300
DB	258 DN-----GFANKTDQFVVAQYQDFGLRPSIATYTKSKAKDVEGIGDXYINIDIGAT 311
QY	301 YFENKMTSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB	312 YFENKMTSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 351
RESULT 3	
ID	Q8X7N5 PRELIMINARY; PRT; 351 AA.
AC	Q8X7N5;
DT	01-MAR-2002 (Tremblrel. 20, Created)
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Outer membrane pore protein E (E.Ic.NmpAB) (Outer membrane pore
DE	protein PhoE).
GN	PHOE OR O302 OR EGS0268.
OS	Escherichia coli O157:H7.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=83334;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC	MEDLINE=21074935; PubMed=11206551;
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA	Welch R.A., Blattner F.R.;
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL	Nature 409:529-533(2001).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=O157:H7 / RMD 0509952;
RC	MEDLINE=21156231; PubMed=11258796;
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli
RT	O157:H7 and genomic comparison with a laboratory strain K-12.";
RL	DNA Res. 8:11-22(2001).
DR	EMBL: AE005202; ANG54566.1; -.
DR	EMBL: AP002551; BAB33691.1; -.
DR	InterPro: IPR003229; OMP_2.
DR	InterPro: IPR001702; Porin_gram-ve.
DR	Pfam: PF00267; Gram-ve_porins; 1.
DR	PRINTS: PR00182; ECOLNEIPORIN.
DR	ProDom: PD000808; OMP_2; 1.
DR	PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW	Complete proteome.
SQ	SEQUENCE 351 AA; 38966 MW; 3E58F4F284B5D42E CRC64;
Query Match	
61.5%; Score 1113; DB 16; Length 351;	


```
Qy 230 NYGTRNATPTINKTNTSGFANKTQDVLLVAQYQDFGLRPSIAYTSKAKDVE-GIGD 288
Db 254 QYTQTNATRVGN-----LGWANKAQNEFVVAQYQDFGLRPSVAYLQSKGKDLNGYQF 308
Qy 289 VDLVNYFEVGATYYFNKNMSTVVDYIIQIDS--DNKLGVSDDTVAVGIVYQF 340
Db 309 QDLKAYVDGATYYFNKNMSTVVDYKINLLDDKEFTRNAGISTDDIVALGLVYQF 363

RESULT 6
O87754
ID O87754 PRELIMINARY; PRT; 374 AA.
AC O87754;
DT 01-NOV-1998 (TremBrel. 08, Created)
DT 01-NOV-1998 (TremBrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
DE OmpK37 porin precursor.
GN OMPK37.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB8;
RX MEDLINE=99235749; PubMed=10217760;
RA Domenech-Sanchez A., Hernandez-Alles S., Martinez-Martinez L.,
RA Benedi V.J., Alberti S.;
RT "Identification and characterization of a new porin gene of Klebsiella
RT pneumoniae: its role in beta-lactam antibiotic resistance.";
RL J. Bacteriol. 181:2726-2732(1999).
DR EMBL; AJ011502; CAA09666.1; -.
DR HSSP; Q48473; 10SM.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PD000808; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Porin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 374 OMPK37 PORIN.
SQ SEQUENCE 374 AA; 41612 MW; B66DF074A1A169B CRC64;

Query Match 57.7%; Score 1044.5; DB 2; Length 374;
Best Local Similarity 58.5%; Pred. No. 1.2e-56;
Matches 214; Conservative 37; Mismatches 76; Indels 39; Gaps 9;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSGYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 22 AEIYNKDGKVDLYGKVDGLHYFSSDSK-----DGDQTYLRFPGKEGTQINDILTYGQ 76
Qy 61 WEYNFQGNSEGAADQTCNKTFLAPAGLYADVGSFDRYGVVYDALGYTDMLPFEGG 120
Db 77 WEYNQANNTETSSQAW--TLAPAGLYKVDGYGSDYGRNYGVLDYVEGTDILPFEGG 134
Qy 121 DT-AYSDOFFYGRVGVATYRNSNFFGLVDGLNFAVQYLGRNEDTA----- 166
Db 135 DSYTYADNFMAGRANGVATYRNSDFGLVEGLNFAVQYQGNNEGQADINVTNNRSSD 194
Qy 167 ---RRNSGDVGGSISYEYEGFI--VGAYGAADRNLQEAQPLGNGKKAQWATGLKYD 221
Db 195 SDVRPDNGDGFGLSYDF-GWISAAAYTSSDRNDQMTNARGDKAEAWTAGLYD 253
Qy 222 ANNIYLAANYGETRNATPTINKTNTSGFANKTQDVLLVAQYQDFGLRPSIAYTSKAK 281
Db 254 ANDIYLATMYSETRNMTPYGN-----DGVANKTQNFVTAQYQDFGLRPAISLYQSKG 308
Qy 282 DVEGIG----VDVLVNYFEVGATYYFNKNMSTVVDYIIQIDSNDKL-----GVGSDTVA 334
Db 309 DLYNNGRYADKDLVYMDVGATYYFNKNMSTVVDYKINLLDGNKDFYEDNGISTDNI 368
```

```
Qy 335 GIVYQF 340
Db 369 GLVYQF 374

RESULT 7
O85030
ID O85030 PRELIMINARY; PRT; 377 AA.
AC O85030;
DT 01-NOV-1998 (TremBrel. 08, Created)
DT 01-NOV-1998 (TremBrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
DE Porin OmpN.
GN OMPN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BE;
RX MEDLINE=98317278; PubMed=9642192;
RA Prilipov A., Phale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
RT "Identification and characterization of two quiescent porin genes,
RT nmpc and ompN, in Escherichia coli BE.";
RL J. Bacteriol. 180:3388-3392(1998).
DR EMBL; AF035618; AAC38644.1; -.
DR HSSP; P02931; IGFN.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PD00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 377 AA; 41236 MW; 7F4D681A2BC10F8C CRC64;

Query Match 57.3%; Score 1036; DB 2; Length 377;
Best Local Similarity 57.2%; Pred. No. 4.1e-56;
Matches 210; Conservative 41; Mismatches 78; Indels 38; Gaps 9;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSGYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 22 AEIYNKDGKVDLYGKVDGLHYFSDNSAK-----DGDQSYARLGPKEGTQINDLTGYGQ 76
Qy 61 WEYNFQGNSEGAADQTCNKTFLAPAGLYADVGSFDRYGVVYDALGYTDMLPFEGG 120
Db 77 WEYNQANNTETSSKNQSW--TLAPAGLYKVDGYGSDYGRNYGVMDIEGWTDMLPFEGG 134
Qy 121 DT-AYSDOFFYGRVGVATYRNSNFFGLVDGLNFAVQYLGRNE-----RDTAR 167
Db 135 DSYTNADNFMAGRANGVATYRNTDFGLVGLNFAVQYQGNNEGASNGEGTNNCRD-VR 193
Qy 168 RNSGDVGGSISYEY-EGFIVGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIY 226
Db 194 HENGSGWGLSTYDILGMSGFASAGYATSSDRNDQVNHHTAAGGDKADAWTAGLYDANNIY 253
Qy 227 LAANYGETRNATPTINKTNTSGFANKTQDVLLVAQYQDFGLRPSIAYTSKAKDVE-- 284
Db 254 LATMYSETRNMTPFGD---SDYAVANKTQNFVTAQYQDFGLRPAVSLMSKGRDLHAA 310
Qy 285 -----GIGDVLVNYFEVGATYYFNKNMSTVVDYIIQIDSNDKL-----GVGSDTVA 333
Db 311 GGADNPAGVDKDLVYADVGATYYFNKNMSTVVDYKINLLDEDDSFYTANGISTDDIVA 370
Qy 334 GIVYQF 340
Db 371 GLVYQF 377

RESULT 8
Q93K39
ID Q93K39 PRELIMINARY; PRT; 372 AA.
```

```
AC Q93K39;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE OmpK36 protein precursor.
GN OmpK36.
GS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=103624;
RA Crowley B., Benedi V.J., Domenech-Sanchez A.;
RT "Porin deficiency and SHV-2 expression results in increased resistance
RT to cephalosporins and carbapenems in Klebsiella pneumoniae.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ344089; CAC50885.1; -
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; UNKNOWN_1.
KW SIGNAL.
FT CHAIN 22 372 OMPK36.
FT CHAIN 22 372 POTENTIAL.
SQ SEQUENCE 372 AA; 40753 MW; 94B84DA54AF97736 CRC64;
Query Match 57.2%; Score 1035.5; DB 2; Length 372;
Best Local Similarity 58.4%; Pred. No. 4.3e-56;
Matches 209; Conservative 38; Mismatches 86; Indels 25; Gaps 6;
QY 1 AEIYNKDGKVDLYGKAVGLHYFSKGENSGYGGNDMTYARLFGKGTQINSLDTGYGQ 60
DB 22 AEIYNKDGKVDLYGKAVGLHYFSKGENSGYGGNDMTYARLFGKGTQINSLDTGYGQ 60
QY 61 WEYNFQGNBSGADQGTGKTRLAFAGLYKADVGSFDYGRNYYGVVYDALGYTMDLPEFG 120
DB 77 WEYNQANNTSSSDQAW--TRLAFAGLKFQDAGSFYGRNYYGVVYDVTWTDVLPEFG 134
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----ARR 168
DB 135 DTYSDFNQLSQRANGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----ARR 194
QY 169 SNGDVGSGSISYE-YEGFIVGAYGAADRTNLQEAQPLNGKKAQEWATGLKYDANNIYL 227
DB 195 QNGDGFGTSLTYDIYDGSISAGFYACSKRNGDQNRDLKGRGDNAETYTGGLYDANNIYL 254
QY 228 AANYGETRNATPIT--NKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTKSKAKDVEG 285
DB 255 ATQYQTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSIATYTKSKAKDVEG 314
QY 286 IGDVLDVNYFEVGATYFENKMNSTYVDYIINQIDSN---KLGVGSDDTVAAGVIYQF 340
DB 315 YGDQDLKYVDVGATYFENKMNSTYVDYKINLLDENDFTRSGISTDDVVALGLVYQF 372
RESULT 9
Q8XE41 ID Q8XE41 PRELIMINARY; PRT; 367 AA.
AC Q8XE41;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein 1b (Ib,c) (Outer membrane protein C OmpC).
GN OmpC OR Z3473 OR ECS3104.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
```

```
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shigenaga H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005453; AAG57350.1; -
DR EMBL; AP002560; BAB36527.1; -
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 367 AA; 40508 MW; 028A0146D462CAA4 CRC64;
Query Match 57.1%; Score 1033; DB 16; Length 367;
Best Local Similarity 58.7%; Pred. No. 6e-56;
Matches 212; Conservative 43; Mismatches 70; Indels 36; Gaps 9;
QY 1 AEIYNKDGKVDLYGKAVGLHYFSKGENSGYGGNDMTYARLFGKGTQINSLDTGYGQ 60
DB 22 AEIYNKDGKVDLYGKAVGLHYFSKGENSGYGGNDMTYARLFGKGTQINSLDTGYGQ 60
QY 61 WEYNFQGNBSGADQGTGKTRLAFAGLYKADVGSFDYGRNYYGVVYDALGYTMDLPEFG 120
DB 77 WEYQIQNSAEN---ENNSWTRVAFAGLKFQDVGSDYGRNYYGVVYDVTWTDVLPEFG 133
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGK-----NERDTARRSN 170
DB 134 DTYSDFNQLSQRANGVATYRNTDFFGLVDGLNFAVQYLGKSGEGMTNNGREALQN 193
QY 171 GDVGSGSISYEYEGFIVGAYGAADRTNLQEAQPL--GNGKKAQEWATGLKYDANNIYL 228
DB 194 GDVGSGSITTYDEYEGFIVGAYGAADRTNLQEAQPL--PLYIGSGDRAETTYTGLKYDANNIYL 252
QY 229 ANYGETRNATPITNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTKSKAKDVEGIG- 287
DB 253 AQYQTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSIATYTKSKAKDVEGIG 307
QY 288 ----DVLVNYFEVGATYFENKMNSTYVDYIINQIDSNKL----GVGSDDTVAAGVIYQ 339
DB 308 RNYDDEDILKYVDVGATYFENKMNSTYVDYKINLLD--DNQFTRDAGINTDNVALGLVYQ 366
QY 340 F 340
DB 367 F 367
RESULT 10
Q9K597 ID Q9K597 PRELIMINARY; PRT; 364 AA.
AC Q9K597;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

DE Outer membrane porin C precursor.
GN OMP_C.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Low A.S., Mackenzie F.M., Gould I.M., Booth I.R.;
RT "Parallel evolution of multi-resistant bacteria in a patient with
RT recurrent septicemia; unique data that support the presence of
RT separate protected environments.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
DR EMBL; AJ295721; CAC01403.1; -.
DR HSP; O52503; IIIV.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; PorIn_gram-ve.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
DR Outer membrane; Porin; Signal; Transmembrane.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 364 AA; 40312 MW; 9B583F2C11344E31 CRC64;
Query Match 56.9%; Score 1029.5; DB 2; Length 364;
Best Local Similarity 58.5%; Pred. No. 9.8e-56;
Matches 209; Conservative 40; Mismatches 77; Indels 31; Gaps 8;
QY 1 AEIYNKDGKVDLYGKAVGLHYFSGKNGSYGNGDMTYARLFGKETQINSLDTGYGQ 60
Db AEIYNKDGKVDLYGKAVGLHYFSGKNGSYGNGDMTYARLFGKETQINSLDTGYGQ 60
22 AEIYNKDGKVDLYGKAVGLHYFSGKNGSYGNGDMTYARLFGKETQINSLDTGYGQ 76
QY 61 WEYFQGNNSGADAQGTNTRFAGLKYADVGSFDFGRNYGVVYDGLGYDMLPFGG 120
Db WEYFQGNNSGADAQGTNTRFAGLKYADVGSFDFGRNYGVVYDGLGYDMLPFGG 120
77 WEYIQGNEPE---SDNSSWTRVAFAGLKQDVGDFGRNYGVVYDVTWTDVLPFGG 133
QY 121 DTAYSDDFVGRGVGATYRNSNFFGLVDGLNFVAQYLGK-----NERDTARRS 169
Db DTAYSDDFVGRGVGATYRNSNFFGLVDGLNFVAQYLGK-----NERDTARRS 169
134 DTVDSDNFMQQRNGGFATYRNTDFGLVDGLDFAVQYQGRNGSAHGEWITNGRDVFEQ 193
QY 170 NGDVGSGSISYEYEGFIVGAYGAADRT-NLQEAQPLGNGKKABQWATGLKYDANNIYLA 228
Db NGDVGSGSISYEYEGFIVGAYGAADRT-NLQEAQPLGNGKKABQWATGLKYDANNIYLA 228
194 NGDVGSGSISYEYEGFIVGAYGAADRT-NLQEAQPLGNGKKABQWATGLKYDANNIYLA 253
QY 229 ANYGETRNATPITNKEPTNTSGFANKTDVLLVAQYDFGLRPSIAVTKSKADV-EGIG 287
Db ANYGETRNATPITNKEPTNTSGFANKTDVLLVAQYDFGLRPSIAVTKSKADV-EGIG 287
254 AQYTQTNATRV-----GSLGWANKAQNFEVAQYDFGLRPSLAYLQSKGNLGRGYD 308
QY 288 DVDLVDFEYGVATYFKNKNSYDYIINQIDSDNKL----GVGSDDTVAVGIVYQF 340
Db DVDLVDFEYGVATYFKNKNSYDYIINQIDSDNKL----GVGSDDTVAVGIVYQF 340
309 DEDILKYVDVGATYFKNKNSYDYIINQIDSDNKL----GVGSDDTVAVGIVYQF 364
RESULT 11
Q9K3E6 PRELIMINARY; PRT; 378 AA.
ID Q9K3E6
AC Q9K3E6
DT 01-OCT-2000 (TtEMBLrel. 15, Created)
DT 01-OCT-2000 (TtEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TtEMBLrel. 21, Last annotation update)
DE Outer membrane protein C precursor.
GN OMP_C.
OS Salmonella enterica subsp. enterica serovar Minnesota.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=70803;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-Sflll1;
RA Zimmermann H., Wassenaar T.M., Laubenheimer-Preusse H., Petry F.,
EA Loos M.;
DR Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: HOMOTRIMER [BY SIMILARITY].
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
    CC      (BY SIMILARITY).
DR EMBL; Y15844; CAB96613.1; -.
DR HSP; O52503; IIIV.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-mot.
DR Pfam; PF00267; Gram-v_porins.1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR ProDom; PD00808; OMP_2; 1.
DR POSITE; PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT SIGNAL          1         21        POTENTIAL.
FT CHAIN           22       378        OUTER MEMBRANE PROTEIN C.
SQ SEQUENCE        378 AA; 41269 MW; 254524EB9EC3849C CRC64;

Query Match              56.9%; Score 1029.5; DB 2; Length 378;
Best Local Similarity   57.0%; Pred.No.1e-55;
Matches 208; Conservative 46; Mismatches 78; Indels 33; Gaps

QY      1 AEIYNKDGKVLDYGKAVGLHYFSKGNGSNYSGGNDMTYARLGFGKETQINSDLTYGQG 60
Db      22 AEIYNKDGKNLIDLFKGYDGLHYS-----DDKSGSDQTMYMRIGFKGETQVNDQLTYGQ 76
QY      61 WEYNFQGNSEGADAOGTNKTRLAFLAGLKAYADVGSFDYGRNYSVVVDALGYTDMLEPFG 120
Db      77 WEYOIQNQTEGSS---DSWIRVAEFLAKFPADAGSEFDYGRNYGVTVTDVTDLPEFG 133
QY      121 DTAISDDFFYGRYGVGATYRNNSFFEGLVGDLNFVAVOYLGN-----ERDTAR---RSNGD 172
Db      134 DTYGADNMFMQRNGVATYRTNTDFEGVLGCLDFALQYQCKNGSVSGENDGSRLLNQGD 193
QY      173 GVGSISIEY-EFGFIVGAY----GAADRNLQAEPALPGKKAEQWATGLKYDVANNIYL 227
Db      194 YGGSLSLYAIGEVSFVGSAITTSKRKTADQNNTADARLYGNDRATVYTGLKYDANNIYL 253
QY      228 AANYGETRNAITIPI-----TNKFTNTSGFANKTQDVLVAOYQDFGLRPSYATYTKSKADOV 283
Db      254 AAQSYQYNATRGTSGNKNKSOSYGFANKAQFEVVAQQYQDFGLRPSVAYLQSXGKDI 313
QY      284 E-----GIGDVDLVNYPEVGATYYFNKNMSTYYDYIIINOIDSND--KLGVGSDDTVAVG 335
Db      314 SNGYASYGQDDIKVYVDVGATYYFNKNMSTYYDYKINKLLDKNDFRDAGINTDDIVALG 373
QY      336 IVYQF 340
Db      374 LVYQF 378

RESULT 12
Q9RH85 PRELIMINARY; PRT; 366 AA.
AC Q9RH85;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN OmpC.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RA Yu S.L., Syu W.J.;
```

Altered outer membrane protein OmpC in hemorrhagic Escherichia coli O157:H7;
Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: AF057355; AAF21761.1; -.
DR HSSP: O52503; 1IIV.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR PRODOM: PD000808; OMP_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
SQ SEQUENCE 366 AA; 40499 MW; 6A4EAD1652565C00 CRC64;

Query Match 56.9%; Score 1028.5; DB 2; Length 366;
Best Local Similarity 58.6%; Pred. No. 1.1e-55;
Matches 211; Conservative 43; Mismatches 71; Indels 35; Caps 9;

QY 1 AEIYNKGNKVDLYGRAVGLHYFSKNGENSYGNGDMTYARLGFKGETQINSDLTGYGQ 60
DB 22 AEIYNKGNKLDLYGKVDGLHYFS-----DDKSVGDQTYMRLGFKGETQVTDQLTGYGQ 76
QY 61 WEYNFOGNNSEGADAOQGNKTRLAFAAGLYADYGSFDYGRNYGVVVDALGYTDMLPFEGG 120
DB 77 WEYIOGNSAEN---ENNSWTRVAFAGLRFQDVGSESDYGRNYGVVVDVTSWTDVLPFEGG 133
QY 121 DTAYSDOFFVGRVGVGATYNSFFGLVDGLNFAVQYLGK-----NERDTARRSNG 171
DB 134 DTYGSDNFQMQRNGFATYNTDFGLVDGLNFAVQYQKNGSVSEGTNNRGALRQNG 193
QY 172 DVGGSISYEYEGFVIGAYGAADRNLQEAQPL--GNGKKAQOMATGLKYDANNIYLA 229
DB 194 DVGGSITDYEGFICAAVSSSKRTDDQNS-PLYICNGDRAETYTGLKYDANNIYLA 252
QY 230 NYGETRATPITNKTNTSGFANKTODVLLVAQYQDFGLRPSIATYTKSKAKDVEGIG-- 287
DB 253 QYTQTNTATRV-----GSLGWAKNAQNFVAQYQDFGLRPSLAYLQSKGNLGVNGR 307
QY 288 ---DVDLVNTEFVGATYFNKNNSTVDYIINOISDNKL-----GVGSDTAVGVYQF 340
DB 308 NYDDEDILKYVDVGVGATYFNKNNSTVDYKINLLD-DNQFTRDAGINTDNVALGLFYQF 366

RESULT 13
Q9RAW3
ID Q9RAW3 PRELIMINARY; PRT; 333 AA.
AC Q9RAW3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Phosphate porin (Fragment).
GN PHOE.
OS Klebsiella pneumoniae (subsp. ozaenae).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_Taxid=574;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1436;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S., Kemp D.J.;
RA "Phylogenetic analysis of Calymmatobacterium granulomatis.";
RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RL Int. J. Syst. Bacteriol. 49:1695-1700(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).
CC EMBL: AF009229; AAD21517.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.

DR PFAM: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PRODOM: PD000808; OMP_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1 1

DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PRODOM: PD000808; OMP_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1 1
SQ SEQUENCE 333 AA; 36839 MW; 58AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;
Best Local Similarity 58.9%; Pred. No. 2.4e-55;
Matches 196; Conservative 41; Mismatches 79; Indels 17; Caps 5;

QY 1 AEIYNKGNKVDLYGRAVGLHYFSKNGENSYGNGDMTYARLGFKGETQINSDLTGYGQ 60
DB 15 AEIYNKGNKLDYGIKAMHYFSYDSK-----DGOYTVRFGIKGETQINDDLTYGR 69
QY 61 WEYNFOGNNSEGADAOQGNKTRLAFAAGLYADYGSFDYGRNYGVVVDALGYTDMLPFEGG 120
DB 70 WESEFSGNKTESDSSQ---KTRLAFAAGLYKNGSPDYGRNLGALYDVEATDMFPEFG 126
QY 121 D-TAYSDOFFVGRVGVGATYNSFFGLVDGLNFAVQYLGKNERDTRARRSNGDVGSGSIS 179
DB 127 DSSAQTDNFWTKRASGLATYRNTDFGLVDGLD/LT/LQYQCKNEGREAKKQNGDVGVTSL 186
QY 180 YEYEG--FGIVGAYGAADRNLQEAQPLGNGKKAQOMATGLKYDANNIYLAANYGETRNA 237
DB 187 YDEGGSDFAVSAAYTSSDRTDQNLARGQSGKAEAWATGLKYDANNIYLATYSETRKM 246
QY 238 TPITNKTNTSGFANKTODVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNTYFEV 297
DB 247 TPI-----SGGFANKAQNFEVAQYQDFGLRPSLGIVLSKCKDIEGVGSEDLVNTIDV 300
QY 298 GATYFNKNNSTVDYIINOISDNKLGVGSD 330
DB 301 GLTYFNKNNNAFVDYKINQLKSDNKLGINDD 333

RESULT 14
Q9RAW2
ID Q9RAW2 PRELIMINARY; PRT; 333 AA.
AC Q9RAW2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Phosphate porin (Fragment).
GN PHOE.
OS Klebsiella pneumoniae subsp. rhinoscleromatis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_Taxid=39831;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC5046;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S., Kemp D.J.;
RA "Phylogenetic evidence for reclassification of Calymmatobacterium granulomatis as Klebsiella granulomatis comb. nov.";
RT Int. J. Syst. Bacteriol. 49:1695-1700(1999).
RL Int. J. Syst. Bacteriol. 49:1695-1700(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).
CC EMBL: AF009230; AAD21518.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PRODOM: PD000808; OMP_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1 1

[illegible]

Search completed: December 18, 2002, 16:21:01
Job time : 22.0203 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 4.94091 Seconds
(without alignments)
1995.696 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377

Sequence: 1 MRGSHHHHGHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	377	100.0	74	22	AA82613
2	279.5	74.1	76	21	AA810562
3	243.5	64.6	76	21	AA810561
4	221	58.6	84	19	AAW43025
5	213	56.5	84	19	AAW43026
6	205	54.4	84	19	AAW43027
7	199	52.8	42	21	AA810558
8	197	52.3	84	19	AAW43028
9	181	48.0	84	19	AAW43024
10	159	42.2	42	21	AA810557

11	124	32.9	42	19	AAW43029	Amino acid sequenc
12	121	32.1	43	21	AA808364	Peptide used to as
13	121	32.1	43	21	AA808383	HIV protease-associ
14	119.5	31.7	103	22	AA870797	Synthetic self-ass
15	118.5	31.4	93	22	AA870796	Immoblised subst
16	116	30.8	94	22	AA86459	Amino acid sequenc
17	115	30.5	43	21	AA808344	Peptide used to as
18	114.5	30.4	43	21	AA808375	Synthetic self-ass
19	113	30.0	43	21	AA808362	Zip5 polypeptide u
20	113	30.0	43	21	AA808366	Recombinant ubiqul
21	111.5	29.6	43	21	AA808386	Human procaspase 8
22	108	28.6	94	22	AA86460	Human caspase-8
23	104	27.6	43	21	AA808378	Recombinant ubiqul
24	103	27.3	113	23	AA883407	Human procaspase 8
25	101.5	26.9	286	21	AA556997	Recombinant ubiqul
26	101.5	26.9	286	21	AA556997	Human caspase-8
27	101	26.8	117	23	AA883405	Recombinant ubiqul
28	97	25.7	291	22	AA882608	Spider recombinant
29	97	25.7	681	22	AA882609	Recombinant Sp 100
30	96.5	25.6	491	18	AAW13373	P.aeruginosa E coli
31	96	25.5	175	21	AA444383	Human procaspase 8
32	95.5	25.3	128	22	AA866825	Human procaspase 8
33	95.5	25.3	128	22	AA866828	Human procaspase 8
34	95	25.2	373	14	AA839486	Human apoAIV mutel
35	94	24.9	168	21	AAV44384	P.aeruginosa H co
36	94	24.9	168	21	AAV44380	P.aeruginosa H co
37	94	24.9	175	21	AAV44382	P.aeruginosa H co
38	92	24.4	145	20	AAW42666	H6FtripB-UB fusio
39	92	24.4	159	19	AAW31552	Collagen binding p
40	92	24.4	493	12	AA813992	P. faiciparum spor
41	91.5	24.3	168	21	AAV44385	P.aeruginosa E coli
42	91.5	24.3	169	21	AAV44381	P.aeruginosa E coli
43	91	24.1	18	21	AA810560	Coiled coil protei
44	91	24.1	379	10	AA890508	Sequence of an epi
45	90.5	24.0	380	10	AA890509	Sequence of an epi

ALIGNMENTS

RESULT 1

AA82613

ID AA82613 standard; Protein; 74 AA.

XX

AC AA82613;

XX

DT 02-OCT-2001 (first entry)

XX

DE Recognin B1 protein.

XX

KW Recognin B1; structural protein; purification; fibre; spinning.

XX

OS Synthetic.

XX

PN W0200153333-Al.

XX

PD 26-JUL-2001.

XX

XX 01-NOV-2000; 2000WO-US30086.

XX

PR 20-JAN-2000; 2000US-0490291.

XX

PA (WELL/) MELLO C M.

PA (ARCI/) ARCDIACONO S.

PA (BUTL/) BUTLER M M.

XX (USSA) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

XX WPI; 2001-483136/52.

DR N-PSDB; AA826305.

XX

PT Recovering structural polypeptides in a biological sample, useful for

PT purifying and spinning spider silks and other structural proteins,
PT comprises treating the sample containing the polypeptides with an acid
PT
XX
PS
PS Claim 2; Page 44; 49pp; English.
XX
XX The present sequence is that of synthetic coiled protein Recogonin
CC Bl. The invention provides methods for purifying structural
CC proteins, including spider silk proteins. Organic acids are used
CC to lyse recombinant cells or other biological samples (such as
CC non-recombinantly derived cells), and enrich the purity and yields
CC of structural proteins by hydrolysing many of the macromolecules
CC while leaving the structural proteins intact. In the present case,
CC Recogonin B1 DNA was cloned into vector pOE-9 for recombinant
CC expression in *Escherichia coli* SG13009pREP4, and recombinant protein
CC was purified from lyophilised pellets using formic acid, acetic acid,
CC propionic acid, butyric acid or valeric acid. Formic and valeric
CC acids, but none of the other organic acids, were able to extract
CC significant quantities of the recombinant protein, and valeric acid
CC was able to extract Recogonin B1 in a relatively pure form. The new
CC method has the following advantages over prior art: it involves
CC fewer steps, requires less time and smaller volumes of reagents,
CC results in better recovery of protein at higher purity (70-99%), is
CC easy to scale up, and fibres can be spun in an environmentally
CC benign solution reducing hazardous waste accumulation and cost.
XX
XX Sequence 74 AA;
SQ
Query Match 100.0%; Score 377; DB 22; Length 74;
Best Local Similarity 100.0%; Pred. No. 1e-32; Mismatches 0; Gaps 0;
Matches 74; Conservative 0; Indels 0; Gaps 0;
QY 1 MRGSHHHHGHGSMASGDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKAKIGDL 60
Db 1 MRGSHHHHGHGSMASGDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKAKIGDL 60
QY 61 NNTSGIRRRPAKLN 74
Db 61 NNTSGIRRRPAKLN 74
RESULT 2
AAB10562
ID AAB10562 standard; protein; 76 AA.
AC AAB10562;
DT 15-DEC-2000 (first entry)
DE Copolymer block protein insert L2-B.
XX
XX Coiled coil protein; copolymer blocker; vulnerary; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW antibiotic; dressing; abrasion; burn.
OS Unidentified.
XX
XX US6090911-A.
PN 18-JUL-2000.
PD 22-OCT-1997; 97US-0956307.
PF 22-OCT-1997; 97US-0956307.
XX
XX (UYMA-) UNIV MASSACHUSETTS.
PA
PI Petka WA, McGrath KP, Tirrell DA;
XX
XX WPI; 2000-542235/49.
DR
XX Synthetic block copolymer with an antibiotic compound, useful for
PT dressing abrasion, burn or non-puncture wound, comprises two

PT alpha-helical protein blocks, random-coil protein block and linker
PT proteins -
XX
PS Example 1; Column 33-36; 53pp; English.
XX
XX This invention describes a novel synthetic block copolymer XYZ comprising
CC two alpha-helical protein blocks X and Z, a water soluble, random-coil
CC protein block Y which links X and Z, and linker proteins for linking the
CC alpha-helical protein blocks to the random-coil protein block. The
CC products of the invention have vulnerary activity. A copolymer produced
CC from a gene that encodes a derivative of AC1_0A copolymer was suspended
CC in water to form a gel, which was then used to treat a wound. The cell
CC binding domain was the integrin ArgGlyAspSer sequence which is known to
CC bind to gp130/IIIa proteins expressed on fibroblast cells necessary for
CC matrix formation for the regeneration of skin. Fibroblasts become
CC entrapped within the gel and thus remain at the site of the wound,
CC serving as a scaffold for the regeneration of tissue surrounding the
CC wound. A synthetic block copolymer together with an antibiotic compound
CC is useful for wound dressing. The copolymer is useful for dressing
CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the
CC polymers of equal size are produced biologically from a single template.
CC The intermolecular binding of alpha-helical blocks are monodisperse which
CC gives a uniform pore size depending on the length of random-coil block.
CC The new copolymers have low molecular weight hence have decreased
CC viscosity solutions or gels under suitable conditions. This sequence
CC represents a copolymer block protein L2-A which is used to illustrate the
CC method of the invention.
XX
XX Sequence 76 AA;
SQ
Query Match 74.1%; Score 279.5; DB 21; Length 76;
Best Local Similarity 89.1%; Pred. No. 2.2e-22;
Matches 57; Conservative 1; Mismatches 1; Indels 5; Gaps 1;
QY 1 MRGSHHHHGHGSMASGDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 1 MRGSHHHHGHGSDDDKDWASGDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKA 60
QY 56 KIGD 59
Db 61 KIGD 64
RESULT 3
AAB10561
ID AAB10561 standard; protein; 76 AA.
AC AAB10561;
DT 15-DEC-2000 (first entry)
DE Copolymer block protein insert L2-A.
XX
XX Coiled coil protein; copolymer blocker; vulnerary; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW antibiotic; dressing; abrasion; burn.
OS Unidentified.
XX
XX US6090911-A.
PN 18-JUL-2000.
PD 22-OCT-1997; 97US-0956307.
PF 22-OCT-1997; 97US-0956307.
XX
XX (UYMA-) UNIV MASSACHUSETTS.
PA
PI Petka WA, McGrath KP, Tirrell DA;
XX
XX WPI; 2000-542235/49.
DR
XX

PT Synthetic block copolymer with an antibiotic compound, useful for
 PT dressing abrasion, burn or non-puncture wound, comprises two
 PT alpha-helical protein blocks, random-coil protein block and linker
 XX proteins -
 XX
 PS Example 1; Column 33-34; 53pp; English.
 XX
 CC This invention describes a novel synthetic block copolymer XYZ comprising
 CC two alpha-helical protein blocks X and Z, a water soluble, random-coil
 CC protein block Y which links X and Z, and linker proteins for linking the
 CC alpha-helical protein blocks to the random-coil protein block. The
 CC products of the invention have vulnerary activity. A copolymer produced
 CC from a gene that encodes a derivative of AC1.0A copolymer was suspended
 CC in water to form a gel, which was then used to treat a wound. The cell
 CC binding domain was the integrin ArgGlyAspSer sequence which is known to
 CC bind to gp130/IL6 proteins expressed on fibroblast cells necessary for
 CC matrix formation for the regeneration of skin. Fibroblasts become
 CC entrapped within the gel and thus remain at the site of the wound,
 CC serving as a scaffold for the regeneration of tissue surrounding the
 CC wound. A synthetic block copolymer together with an antibiotic compound
 CC is useful for wound dressing. The copolymer is useful for dressing
 CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the
 CC polymers of equal size are produced biologically from a single template.
 CC The intermolecular binding of alpha-helical blocks are monodisperse which
 CC gives a uniform pore size depending on the length of random-coil block.
 CC The new copolymers have low molecular weight hence have decreased
 CC viscosity solutions or gels under suitable conditions. This sequence
 CC represents a copolymer block protein L2-A which is used to illustrate the
 CC method of the invention.
 XX

SQ Sequence 76 AA;

Query Match 64.6%; Score 243.5; DB 21; Length 76;

Best Local Similarity 75.0%; Pred. No. 1.4e-16;

Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;

QY 1 MRGSHHHHHHGS-----MASGDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKA 55

DB 1 MRGSHHHHHHGSDDDDKWSGDLNEVAQLEREVRSLEDEAAELEQKVRSLKNEIEDLKA 60

QY 56 KIGD 59

DB 61 EIGD 64

RESULT 4

ID AAW43025 standard; Peptide; 84 AA.

XX AAW43025;

XX 08-MAY-1998 (first entry)

XX Artificial recognition sequence 2.

XX Recognition sequence; self-assembling; protein-based structural material;
 KW spontaneous heterodimerisation; coiled-coil alpha helix;
 KW selectively permeable membrane; coated fabric.

XX Synthetic.

XX US5712366-A.

XX 27-JAN-1998.

XX 25-MAY-1995; 95US-0452592.

XX 25-MAY-1993; 93US-0068948.

XX (USSA) US SEC OF ARMY.

XX Kaplan DL, McGrath KP;

XX

DR WPI; 1998-120029/11.

XX Production of peptide heterodimer(s) - in the fabrication of
 PT self-assembling protein-based structural material(s)

XX Disclosure; Fig 4B; 17pp; English.

XX
 CC AAW43024-28 represent recognition sequences obtained from the library of
 CC recognition sequences coded for by AAV04943. This DNA sequence is
 CC obtained by reverse translation. The codons were chosen to maximise
 CC expression in Escherichia coli, and to introduce useful restriction for
 CC subsequent genetic manipulations. The DNA sequence use a "mixed site"
 CC approach at the first base of the codons for amino acids positions 5 and
 CC 7 of the heptad (see below). The formation of a self-assembling
 CC structural material is mediated by the artificial recognition sequences
 CC (e.g. present sequence). A method of producing a self-assembling
 CC protein-based structural material comprises mixing a multiplicity of
 CC 2 different artificial peptide sequences, designed so as to
 CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
 CC alpha helical secondary structure having at least 2 heptad units. The
 CC artificial peptides spontaneously heterodimerise into a self-assembled
 CC protein-based structural material. The method is used to produce
 CC nanoscale structural materials designed for specific functions e.g.
 CC membranes, fibres, absorbants, reactive materials, etc. A specific
 CC application is in the fabrication of selectively permeable membranes for
 CC protective gloves and coated fabrics for fuel handlers and other
 CC hazardous materials. Materials for medical implants, wound healants and
 CC other medical treatments would also be feasible using this technology.
 XX

SQ Sequence 84 AA;

Query Match 58.6%; Score 221; DB 19; Length 84;

Best Local Similarity 97.9%; Pred. No. 3.9e-16;

Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62

DB 2 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLKN 48

RESULT 5

AAW43026

ID AAW43026 standard; Peptide; 84 AA.

XX AAW43026;

XX 08-MAY-1998 (first entry)

XX Artificial recognition sequence 3.

XX Recognition sequence; self-assembling; protein-based structural material;
 KW spontaneous heterodimerisation; coiled-coil alpha helix;
 KW selectively permeable membrane; coated fabric.

XX Synthetic.

XX US5712366-A.

XX 27-JAN-1998.

XX 25-MAY-1995; 95US-0452592.

XX 25-MAY-1993; 93US-0068948.

XX (USSA) US SEC OF ARMY.

XX Kaplan DL, McGrath KP;

XX WPI; 1998-120029/11.

XX Production of peptide heterodimer(s) - in the fabrication of
 PT self-assembling protein-based structural material(s)

PS Disclosure; Fig 4C; 17pp; English.

XX

CC AAW43024-28 represent recognition sequences obtained from the library of

CC recognition sequences coded for by AAW04943. This DNA sequence is

CC obtained by reverse translation. The codons were chosen to maximise

CC expression in Escherichia coli, and to introduce useful restriction for

CC subsequent genetic manipulations. The DNA sequence use a "mixed site"

CC approach at the first base of the codons for amino acids positions 5 and

CC 7 of the heptad (see below). The formation of a self-assembling

CC structural material is mediated by the artificial recognition sequences

CC (e.g. present sequence). A method of producing a self-assembling

CC protein-based structural material comprises mixing a multiplicity of

CC 2 different artificial peptide sequences, designed so as to

CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil

CC alpha helical secondary structure having at least 2 heptad units. The

CC artificial peptides spontaneously heterodimerise into a self-assembled

CC protein-based structural material. The method is used to produce

CC nanoscale structural materials designed for specific functions e.g.

CC membranes, fibres, absorbants, reactive materials, etc. A specific

CC application is in the fabrication of selectively permeable membranes for

CC protective gloves and coated fabrics for fuel handlers and other

CC hazardous materials. Materials for medical implants, wound healants and

CC other medical treatments would also be feasible using this technology.

XX

XX Sequence 84 AA:

XX

```

Query Match      56.5%; Score 213; DB 19; Length 84;
Best Local Similarity 93.6%; Pred. No. 2.7e-15;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDKNKVAQLKKRKYRSLUKDKAAELKQHVSRLENEIEDLKAKIGDLNN 62
      |||||:::|:|||||:::|:|||||:::|:|||||:::|:|||||:::|:
Db 2 GDKNKVAQLKREYRSLUKDKAAELKQHVSRLENEIEDLKAKIGDLKN 48

```

RESULT 6	
AAW43027	
ID	AAW43027 standard; Peptide; 84 AA.
XX	
XX	
AC	AAW43027;
XX	
XX	
DT	08-MAY-1998 (first entry)
XX	
XX	
DE	Artificial recognition sequence 4.
XX	
XX	
KW	Recognition sequence; self-assembling; protein-based structural material;
KW	spontaneous heterodimerisation; coiled-coil alpha helix;
KW	selectively permeable membrane; coated fabric.

OS	Synthetic.
XX	
PN	US5712366-A.
XX	
XX	27-JAN-1998.
PD	
XX	
XX	25-MAY-1995; 95US-0452592.
PF	
XX	
XX	25-MAY-1993; 93US-0068948.
PR	
XX	
XX	(USSA) US SEC OF ARMY.
PA	
XX	
XX	Kaplan DL, McGrath KP;
PI	
XX	
XX	WPI; 1998-120029/11.
DR	
XX	
XX	Production of peptide hetero:dimer(s) - in the fabrication of
PT	self-assembling protein-based structural material(s)
PT	

Disclosure: Fig 4D; 17pp: English.

expression in *Escherichia coli*, and to introduce useful restriction for subsequent genetic manipulations. The DNA sequence use a "mixed site" approach at the first base of the codons for amino acids positions 5 and 7 of the hepad (see below). The formation of a self-assembling structural material is mediated by the artificial recognition sequences (e.g. present sequence). A method of producing a self-assembling protein-based structural material comprises mixing a multiplicity of 2 different artificial peptide sequences, designed so as to spontaneously heterodimerise. Both artificial peptides have a coiled-coil alpha helical secondary structure having at least 2 heptad units. The artificial peptides spontaneously heterodimerise into a self-assembled protein-based structural material. The method is used to produce nanoscale structural materials designed for specific functions e.g. membranes, fibres, absorbents, reactive materials, etc. A specific application is in the fabrication of selectively permeable membranes for protective gloves and coated fabrics for fuel handlers and other hazardous materials. Materials for medical implants, wound healants and other medical treatments would also be feasible using this technology.

XX Sequence 84 AA;

SQ

Query Match	54.4%	Score 205	DB 19	Length 84
Best Local Similarity	89.4%	Pred. No. 1.9e-14		
Matches 42	Conservative	4	Mismatches 1	Indels 0
				Gaps 0

Q7	16	GDLKNKVAQLKKRVRSLSKDKAAELKQVSRLENEIEDLKAKTGD	LN	62
		: : : : : : :		
D8	2	GDLENAVQLREVRSLKDKAAELKQVSRLENEIEDLKAKTGD	LE	48
		: : : : : :		

RESULT 7
AAB10558
ID AAB10558 standard: protein: 42 AA.

AC	AAB10558;
XX	
XX	
DT	15-DEC-2000 (first entry)
XX	
XX	
DE	Coiled coil protein B1.
XX	
KW	Coiled coil protein; copolymer blocker; vulnery; wound treatment;
KW	integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW	antibiotic; dressing; abrasion; burn; B1.

OS	Unidentified.
XX	
PN	US6090911-A.
XX	
PD	18-JUL-2000.

xx
xx PF 22-OCT-1997; 97US-0956307.
xx
xx PR 22-OCT-1997; 97US-0956307.
xx
xx (UYMA-) UNIV MASSACHUSETTS.
xx PA
xx Petka WA, McGrath KP, Tirrell DA;
xx PI
xx WPI: 2000-542235/49.
xx DR

XX Synthetic block copolymer with an antibiotic compound, useful for
PS dressing abrasion, burn or non-puncture wound, comprises two
PT alpha-helical protein blocks, random-coil protein block and linker
PT proteins -
XX
PS Example 1; Column 25-26; 53pp; English.

This invention describes a novel synthetic block copolymer XYZ comprising two alpha-helical protein blocks X and Z, a water soluble, random-coil protein block Y which links X and Z, and linker proteins for linking the alpha-helical protein blocks to the random-coil protein block. The products of the invention have vulnery activity. A copolymer produced from a gene that encodes a derivative of AC 1 0A copolymer was suspended

PT Production of peptide heterodimer(s) - in the fabrication of
PT self-assembling protein-based structural material(s)

XX Claim 2; Columns 13-14; 17pp; English.

XX The present sequence is used to obtain an entire library of recognition
CC sequences. The DNA sequence encoding the present sequence is obtained by
CC reverse translation. The codons were chosen to maximise expression in
CC Escherichia coli, and to introduce useful restriction for subsequent
CC genetic manipulations. The DNA sequence use a "mixed site" approach at
CC the first base of the codons for amino acids positions 5 and 7 of the
CC heptad (see below). The formation of a self-assembling structural
CC material is mediated by the artificial recognition sequences obtained
CC from the present sequence. A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of
CC 2 different artificial peptide sequences, designed so as to spontaneously
CC heterodimerise. Both artificial peptides have a coiled-coil alpha helical
CC secondary structure having at least 2 heptad units. The artificial
CC peptides spontaneously heterodimerise into a self-assembled protein-based
CC structural material. The method is used to produce nanoscale structural
CC materials designed for specific functions e.g. membranes, fibres, the
CC absorbants, reactive materials, etc. A specific application is in the
CC fabrication of selectively permeable membranes for protective gloves and
CC coated fabrics for fuel handlers and other hazardous materials. Materials
CC for medical implants, wound healants and other medical treatments would
CC also be feasible using this technology.

XX Sequence 42 AA;

Query Match 32.9%; Score 124; DB 19; Length 42;

Best Local Similarity 72.5%; Pred. No. 3.3e-06;

Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 16 GDLNKKVAQLKRVSLKDKAAELKQVSRLENIEDLKA 55

DB 2 GDLNKKVAQLKRVSLKDKAAELKQVSRLENIEDLKA 41

RESULT 12

AAB08364

ID AAB08364 standard; peptide; 43 AA.

XX AC AAB08364;

DT 20-DEC-2000 (first entry)

XX Amino acid sequence of a coiled-coil peptide.

DE Binding partner; protein modification; post-translational modification;

KW modulator; coiled-coil structure.

XX Unidentified.

OS Key Location/Qualifiers

FH Cleavage-site 24

FT /note= "thrombin cleavage site"

FT Misc-difference 39 /note= "fluorophore attachment site"

XX WO200050902-A2.

XX 31-AUG-2000.

XX 25-FEB-1999; 2000WO-GB00669.

XX 25-FEB-1999; 99GB-0004398.

XX (FLUO-) FLUORESCENCE LTD.

XX Colyer J, Craig RK, Maschio A, Mezna M;

XX WPI; 2000-572119/53.

PT High throughput assay for monitoring modification of polypeptides and
PT modulation of the modifications -
XX Disclosure; Page 61; 128pp; English.

XX The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC state of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The
CC polypeptides, support and methods can be used to analyse a sample to
CC determine if modification of a polypeptide is taking place and to
CC identify modulators of the modification. This is useful for monitoring
CC the post-translational modification of proteins. AAB08364-65 represent
CC binding partners which have coiled-coil structures, and may be used in
CC the method of the invention to assay for thrombin activity.

XX Sequence 43 AA;

Query Match 32.1%; Score 121; DB 21; Length 43;

Best Local Similarity 46.7%; Pred. No. 7e-06;

Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

QY 5 HHHHHGSMASGDLKKNKVAQLKRVSLKDKAAELKQVSRLENE 49

DB 1 HHHHHGGGIA--QLEQETIAQLEQENRQLEQETIAQLEQETIAKLEQE 43

RESULT 13

AAB08383

ID AAB08383 standard; peptide; 43 AA.

XX AC AAB08383;

DT 20-DEC-2000 (first entry)

XX Peptide used to assay for tobacco etch virus protease activity.

DE Binding partner; protein modification; post-translational modification;

KW modulator; coiled-coil structure.

XX Unidentified.

XX WO200050902-A2.

XX 31-AUG-2000.

XX 25-FEB-2000; 2000WO-GB00669.

XX 25-FEB-1999; 99GB-0004398.

XX (FLUO-) FLUORESCENCE LTD.

XX Colyer J, Craig RK, Maschio A, Mezna M;

XX WPI; 2000-572119/53.

XX High throughput assay for monitoring modification of polypeptides and

PT modulation of the modifications -

XX Example 9; Page 80; 128pp; English.

XX The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC state of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 ; Search time 1.70376 Seconds
(without alignments)
1277.936 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377

Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	279.5	74.1	76	3	US-08-956-307B-18
2	243.5	64.6	76	3	US-08-956-307B-17
3	221	58.6	84	1	US-08-452-592B-7
4	213	56.5	84	1	US-08-452-592B-8
5	205	54.4	84	1	US-08-452-592B-9
6	199	52.8	42	3	US-08-956-307B-2
7	197	52.3	84	1	US-08-452-592B-10
8	181	48.0	84	1	US-08-452-592B-6
9	159	42.2	42	3	US-08-956-307B-1
10	124	32.9	42	1	US-08-452-592B-11
11	101.5	26.9	286	4	US-09-360-017-1
12	96	25.5	175	4	US-09-329-884-18
13	94	24.9	168	4	US-09-329-884-20
14	94	24.9	168	4	US-09-329-884-12
15	94	24.9	175	4	US-09-329-884-16
16	92	24.4	159	4	US-08-856-253-2
17	91.5	24.3	168	4	US-09-329-884-22
18	91.5	24.3	169	4	US-09-329-884-14
19	91	24.1	18	3	US-08-956-307B-4
20	90	23.9	345	4	US-08-856-253-7
21	87	23.1	14	3	US-08-956-307B-3
22	82.5	21.9	110	1	US-08-434-705B-15
23	82.5	21.9	110	2	US-09-086-201-15
24	82	21.8	18	1	US-08-578-649-14
25	81	21.5	42	5	PCT-US94-06655-9
26	81	21.5	48	1	US-08-457-245-21
27	81	21.5	84	4	US-09-299-495F-11

28	81	21.5	126	1	US-08-268-348A-12	Sequence 12, Appl
29	81	21.5	133	1	US-08-268-348A-8	Sequence 8, Appl
30	81	21.5	133	1	US-08-268-348A-10	Sequence 10, Appl
31	80.5	21.4	202	4	US-09-134-001C-5383	Sequence 5383, Ap
32	80.5	21.4	1587	4	US-09-000-094-46	Sequence 46, Appl
33	79.5	21.1	31	2	US-08-662-227-37	Sequence 37, Appl
34	79.5	21.1	31	4	US-09-017-947-37	Sequence 37, Appl
35	79	21.0	32	3	US-08-737-336-1	Sequence 1, Appl
36	79	21.0	39	2	US-08-679-865-39	Sequence 39, Appl
37	79	21.0	39	2	US-08-680-876-39	Sequence 39, Appl
38	79	21.0	39	4	US-09-263-975-39	Sequence 39, Appl
39	79	21.0	69	2	US-08-687-865A-21	Sequence 21, Appl
40	79	21.0	69	4	US-09-043-711-21	Sequence 21, Appl
41	79	21.0	88	2	US-08-690-011A-11	Sequence 11, Appl
42	79	21.0	178	4	US-09-183-841-2	Sequence 2, Appl
43	79	21.0	211	4	US-08-856-253-4	Sequence 4, Appl
44	79	21.0	254	4	US-09-004-731-89	Sequence 89, Appl
45	79	21.0	254	4	US-09-004-731-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1

US-08-956-307B-18

; Sequence 18, Application US/08956307B

; Patent No. 6090911

; GENERAL INFORMATION:

; APPLICANT: Petka, Wendy A.

; APPLICANT: Tirrell, David A.

; APPLICANT: Kevin P. McGrath

; TITLE OF INVENTION: REVERSIBLE HYDROGELS

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS: 24

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/956,307B

; FILING DATE: 22-OCT-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 07880/033001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 76 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-956-307B-18

Query Match 74.1%; Score 279.5; DB 3; Length 76;

Best Local Similarity 89.1%; Pred. No. 2.3e-23;

Matches 57; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 1 MRGSHHHHHGSS-----MASGDLKNKVAOLKRVSLKDKAAELKQEVSRLENEEDLKA 55

|||||

```
Db 1 MRGSHHHHGGDDDDKASGDLKKNVAQLKRVSLKDKAAELKQEVSLKNEIEDLKA 60
QY 56 KIGD 59
    ||||
Db 61 KIGD 64

RESULT 2
US-08-956-307B-17
; Sequence 17, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Pecka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-17

Query Match 64.6%; Score 243.5; DB 3; Length 76;
Best Local Similarity 75.0%; Pred. No. 1.7e-19;
Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;

QY 1 MRGSHHHHGG-----MASGDLKKNVAQLKRVSLKDKAAELKQEVSLKNEIEDLKA 55
    |||||
Db 1 MRGSHHHHGGDDDDKASGDLKKNVAQLKRVSLKNEIEDLKA 60

QY 56 KIGD 59
    ||||
Db 61 EIGD 64

RESULT 3
US-08-452-592B-7
; Sequence 7, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
```

```
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-452-592B-7

Query Match 58.6%; Score 221; DB 1; Length 84;
Best Local Similarity 97.9%; Pred. No. 4.9e-17;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKKNVAQLKRVSLKDKAAELKQEVSLKNEIEDLKAIGDLNN 62
    |||||
Db 2 GDLKKNVAQLKRVSLKDKAAELKQEVSLKNEIEDLKAIGDLKN 48

RESULT 4
US-08-452-592B-8
; Sequence 8, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
```

```
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/068,948
/ FILING DATE: 25 May 1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Donahue, Richard J.
/ NAME: Lanning, John H.
/ NAME: Ranucci, Vincent J.
/ REGISTRATION NUMBER: 22,062 (Donahue)
/ REGISTRATION NUMBER: 34,857 (Lanning)
/ REGISTRATION NUMBER: 29,579 (Ranucci)
/ REFERENCE/DOCKET NUMBER: NA-1096D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 508-233-4510
/ TELEFAX: 508-233-5167
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-452-592B-8

Query Match 56.5%; Score 213; DB 1; Length 84;
Best Local Similarity 93.6%; Pred. No. 3.6e-16;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
Db 2 GDLKNKVAQLEREVSLKDKAAELKQEVSRLENEIEDLKAKIGDLKN 48

RESULT 5
US-08-452-592B-9
/ Sequence 9, Application US/08452592B
/ Patent No. 5712366
/ GENERAL INFORMATION:
/ APPLICANT: McGrath, Kevin P.
/ APPLICANT: Kaplan, David L.
/ TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
/ TITLE OF INVENTION: Self-Assembling Proteins
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: U.S. Army Soldier Systems Command
/ STREET: AMSCC-CC (Patent Counsel)
/ CITY: Natick
/ STATE: Massachusetts
/ COUNTRY: United States of America
/ ZIP: 01760-5035
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: DOS 6.22
/ SOFTWARE: WordPerfect Version 6.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/452,592B
/ FILING DATE: 25 May 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/068,948
/ FILING DATE: 25 May 1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Donahue, Richard J.
/ NAME: Lanning, John H.
/ NAME: Ranucci, Vincent J.
/ REGISTRATION NUMBER: 22,062 (Donahue)
/ REGISTRATION NUMBER: 34,857 (Lanning)
/ REGISTRATION NUMBER: 29,579 (Ranucci)
/ REFERENCE/DOCKET NUMBER: NA-1096D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 508-233-4510
/ TELEFAX: 508-233-5167
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
```

```
/
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-452-592B-9

Query Match 54.4%; Score 205; DB 1; Length 84;
Best Local Similarity 89.4%; Pred. No. 2.6e-15;
Matches 42; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
Db 2 GDLNEVAQLEREVSLKDKAAELKQEVSRLENEIEDLKAKIGDLN 48

RESULT 6
US-08-956-307B-2
/ Sequence 2, Application US/08956307B
/ Patent No. 6090911
/ GENERAL INFORMATION:
/ APPLICANT: Petka, Wendy A.
/ APPLICANT: Tirrell, David A.
/ APPLICANT: Kevin P. McGrath
/ TITLE OF INVENTION: REVERSIBLE HYDROGELS
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: FastSeq for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/956,307B
/ FILING DATE: 22-OCT-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fasse, J. Peter
/ REGISTRATION NUMBER: 32,983
/ REFERENCE/DOCKET NUMBER: 07880/033001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/542-5070
/ TELEFAX: 617/542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 42 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-08-956-307B-2

Query Match 52.8%; Score 199; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 5e-15;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 SGDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 56
Db 1 SGDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 42

RESULT 7
US-08-452-592B-10
/ Sequence 10, Application US/08452592B
/ Patent No. 5712366
/ GENERAL INFORMATION:
/ APPLICANT: McGrath, Kevin P.
/ APPLICANT: Kaplan, David L.
```

```
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; TITLE OF INVENTION: Self-Assembling Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-5167
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-452-592B-10

Query Match 52.3%; Score 197; DB 1; Length 84;
Best Local Similarity 85.1%; Pred. No. 1.9e-14;
Matches 40; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GDLENEVAQLEREVRSLEDEAAELKQEVSRLENEIEDLKAKIGDLN 48

RESULT 8
US-08-452-592B-6
; Sequence 6, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; TITLE OF INVENTION: Self-Assembling Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
```

```
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-452-592B-6

Query Match 48.0%; Score 181; DB 1; Length 84;
Best Local Similarity 76.6%; Pred. No. 9.7e-13;
Matches 36; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2 GDLENEVAQLEREVRSLEDEAAELKQEVSRLENEIEDLKAKIGDLN 48

RESULT 9
US-08-956-307B-1
; Sequence 1, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:21:10 ; Search time 1.02236 Seconds
(without alignments)
1213.493 Million cell updates/sec

Title: US-09-490-291-11
Perfect score: 377
Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	104	27.6	131	9	US-09-982-992A-2
2	101.5	26.9	286	10	US-09-862-915-1
3	96	25.5	175	10	US-09-865-159-18
4	95.5	25.3	128	10	US-09-742-373-9
5	94	24.9	168	10	US-09-865-159-20
6	94	24.9	169	10	US-09-865-159-12
7	94	24.9	175	10	US-09-865-159-16
8	93.5	24.8	323	9	US-09-987-107-58
9	92	24.4	159	10	US-09-813-820-2
10	92	24.4	336	9	US-09-987-107-44
11	92	24.4	337	9	US-09-987-107-46
12	91.5	24.3	168	10	US-09-865-159-22
13	91.5	24.3	169	10	US-09-865-159-14
14	91.5	24.3	324	12	US-10-007-805-551
15	90	23.9	273	9	US-09-987-107-50
16	90	23.9	345	10	US-09-813-820-7
17	89.5	23.7	323	9	US-09-987-107-56
18	88	23.3	324	9	US-09-987-107-62
19	88	23.3	324	9	US-09-987-107-64

20	87.5	23.2	323	9	US-09-987-107-60
21	87	23.1	265	10	US-09-732-091-44
22	83.5	22.1	316	9	US-09-987-107-48
23	83	22.0	504	10	US-09-732-091-42
24	82	21.8	92	10	US-09-833-747A-2
25	82	21.8	324	9	US-09-987-107-66
26	81	21.5	316	9	US-09-987-107-54
27	80	21.2	104	9	US-05-870-759-90
28	79.5	21.1	31	10	US-09-925-442-37
29	79	21.0	39	10	US-09-884-681-39
30	79	21.0	211	10	US-09-813-820-4
31	78	20.7	139	10	US-09-813-820-8
32	78	20.7	439	9	US-10-115-984-2
33	78	20.7	512	10	US-09-813-820-6
34	77	20.4	559	10	US-09-981-649A-28
35	76.5	20.3	14	9	US-09-784-199-9
36	76.5	20.3	193	9	US-09-991-496-119
37	76.5	20.3	193	10	US-09-874-923-119
38	76	20.2	326	9	US-09-991-211-1
39	75	19.9	30	10	US-09-754-826-3
40	74.5	19.8	261	9	US-09-987-107-52
41	74.5	19.8	344	9	US-09-987-107-68
42	74.5	19.8	716	10	US-09-845-157-2
43	74.5	19.8	879	9	US-10-108-605-217
44	74	19.6	17	9	US-09-051-013-6
45	74	19.6	20	9	US-09-051-013-3

ALIGNMENTS

RESULT 1

US-09-982-992A-2
; Sequence 2, Application US/09982992A
; Patent No. US20020164337A1

; GENERAL INFORMATION:

; APPLICANT: PATTI, Joseph M. et al.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO THE MAP PROTEIN AND METHOD OF USE IN

; TITLE OF INVENTION: AND PREVENTING INFECTIONS

; FILE REFERENCE: P06922US02/BAS

; CURRENT APPLICATION NUMBER: US/09/982.992A

; CURRENT FILING DATE: 2001-10-22

; PRIOR APPLICATION NUMBER: 60/277,287

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/241,832

; PRIOR FILING DATE: 2000-10-20

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 131

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-982-992A-2

Query Match 27.6% Score 104; DB 9; Length 131;
Best Local Similarity 30.7% Pred. No. 0.0029;
Matches 31; Conservative 9; Mismatches 27; Indels 34; Gaps 3;

Qy 1 MRGSHHHHHGSGIPYITVNGTSQNLSSLTFNKNQOISYKDIENKVKSVLYFNRGISD 60
|||||
Db 1 MRGSHHHHHGSGIPYITVNGTSQNLSSLTFNKNQOISYKDIENKVKSVLYFNRGISD 60
|||||

Qy 35 -----AAELKQEVSRLENEIEDLKAKI--GDLNNTSGIR 67
|||
Db 61 IDRLSKQAKYTVHFNGTKRKRVLDKAGIHTADLINTSDIK 101
|||

RESULT 2

US-09-862-915-1
; Sequence 1, Application US/09862915
; Patent No. US20020045205A1

; GENERAL INFORMATION:

; APPLICANT: Heinrichson, Robert I.

```
; APPLICANT: Tomasselli, Alfredo G.
; TITLE OF INVENTION: Method for Autoactivation of Procaspase 8
; FILE REFERENCE: Docket No. US2002045205A1 6172
; CURRENT APPLICATION NUMBER: US/09/862,915
; PRIOR FILING DATE: 2001-05-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-862-915-1

Query Match      26.9%; Score 101.5; DB 10; Length 286;
Best Local Similarity 34.6%; Pred. No. 0.011;
Matches 27; Conservative 8; Mismatches 20; Indels 23; Gaps 3;

QY 1 MRGSHHHHHGSMASGD-----LKNKVAQLKKRVSL-----KDKAAELKQEVSR 45
Db 1 MRGSHHHHHGSMTISDPREODSESQTLDKVYQMKSKPRGYCLIIINNNHFAKAREKVPK 60

QY 46 LENEIEDLKAIGDLNNT 63
Db 61 LHS-----IRDRNGT 70

RESULT 3
US-09-865-159-18
; Sequence 18, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-865-159-18

Query Match      25.5%; Score 96; DB 10; Length 175;
Best Local Similarity 28.6%; Pred. No. 0.021;
Matches 22; Conservative 15; Mismatches 24; Indels 16; Gaps 1;

QY 5 HHHHHGSMASGDLKNKVAQLKKRVSLKDKAAELKQEVSRLENE----- 49
Db 4 HHHHGGGGEVSALEKEVSALEKEVSALEKEVSALEKGGGGBFARAQLSERMT 63

QY 50 -IEDLKAIGDLNNTSG 65
Db 64 LASGLTKVSDIFSQDG 80

RESULT 4
US-09-742-373-9
; Sequence 9, Application US/09742373
; Patent No. US20020052471A1
; GENERAL INFORMATION:
; APPLICANT: Althaus, Harald
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: Human Procalcitonin and the Preparation and Use Thereof
; FILE REFERENCE: 05552.1445-00
; CURRENT APPLICATION NUMBER: US/09/742,373
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 19962434.8
```

```
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 10016278.9
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 10027954.6
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Protein, human
; OTHER INFORMATION: procalcitonin
; US-09-742-373-9

Query Match      25.3%; Score 95.5; DB 10; Length 128;
Best Local Similarity 30.8%; Pred. No. 0.017;
Matches 24; Conservative 18; Mismatches 17; Indels 19; Gaps 3;

QY 1 MRGSHHHHHHGS-----MASGDLKNKVAQLKKRVSLKDKAAELKQEV 44
Db 1 MRGSHHHHHHGSAPFRSALESSPADPATUSEEARLRLAALVQDY--VOMKASELEQEQE 58

QY 45 RLENEIEDLKAIGDLN 61
Db 59 REGSSLDSPRSKRCGNLS 76

RESULT 5
US-09-865-159-20
; Sequence 20, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-865-159-20

Query Match      24.9%; Score 94; DB 10; Length 168;
Best Local Similarity 40.0%; Pred. No. 0.03;
Matches 20; Conservative 8; Mismatches 16; Indels 6; Gaps 1;

QY 6 HHHHHGSMASGDLKNKVAQLKKRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 4 HHHHHGSG-----GGEIEALKAEIEALKAEIEALKAEIEALKAE 47

RESULT 6
US-09-865-159-12
; Sequence 12, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
```



```

; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 551
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-551

Query Match      24.3%; Score 91.5; DB 12; Length 324;
Best Local Similarity 32.4%; Pred. No. 0.1;
Matches 24; Conservative 11; Mismatches 24; Indels 15; Gaps 2;

QY 5 HHHHHHGSMA-----SGDLKNKVAQLKRRKRSKDKAAELKOEVSRLN-----E 49
   |||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 HHHHHGTRALQCEVSTHENENYLLHENCMLKKEIAMLKLEIATLKHQYQEKENKYFED 62

QY 50 IEDLKAKIGDLNNT 63
   | : | | | : | |
Db 63 IKILKEKNAELQMT 76

RESULT 15
US-09-987-107-50
; Sequence 50, Application US/09987107
; Patent No. US20020156007A1
; GENERAL INFORMATION:
; APPLICANT: GRAVERSEN, Jonas
; APPLICANT: MOESTRUP, Soren
; TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES
; FILE REFERENCE: GRAVERSEN1A
; CURRENT APPLICATION NUMBER: US/09/987,107
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/264,022
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: DK PA2001 00057
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: DK PA2000 01682
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pT7H6 Trip-A-Apo A-1-del 43 - AmpR plasmid
US-09-987-107-50

Query Match      23.9%; Score 90; DB 9; Length 273;
Best Local Similarity 29.2%; Pred. No. 0.12;
Matches 26; Conservative 11; Mismatches 34; Indels 18; Gaps 3;

QY 3 GSHHHHHHGSMA-----GDLKNKVAQLKRRK-----RSLKAAELKOEVSRLN 48
   |||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 GSHHHHHHSIQGRSPGTEPTQPKKIVNAKKDVNTKMFEEKLSRLDTLAQEVALLKE 61

QY 49 ----EIEDLKAKIGDLNNTSGIRRPAAKL 73
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 62 QQALQTVSLKGSCLKLLDNWDSVTSTFSKL 90

Search completed: December 18, 2002, 16:39:47
Job time : 17.0223 secs
```

This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 2.18649 Seconds
(without alignments)
3253.588 Million cell updates/sec

Title: US-09-490-291-11
Perfect score: 377
Sequence: 1 MRCGSHHHHGHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	94.5	25.1	1186	2 G69708	chromosome segrega
2	89	23.6	315	2 JC7572	somite Waf1 protei
3	88	23.3	259	2 A70359	hydrogenase expres
4	86	22.8	155	2 I50167	B-G antigen - chic
5	85.5	22.7	323	2 I49529	transcription fact
6	85	22.5	321	2 D39371	Ig V-region-like B
7	85	22.5	880	2 F75103	conserved hypothet
8	85	22.5	979	2 JQ0894	p115 protein - Myc
9	84.5	22.4	281	2 F75216	hypothetical prote
10	84	22.3	1199	2 T29145	hypothetical prote
11	82.5	21.9	397	2 F90182	hypothetical prote
12	82	21.8	173	2 JC5610	troponin I - sea s
13	82	21.8	311	2 A56235	transcription acti
14	82	21.8	766	2 S37894	hypothetical prote
15	82	21.8	1188	2 G83960	chromosome segrega
16	81.5	21.6	212	2 H69402	hypothetical prote
17	81	21.5	166	2 S73342	hypothetical prote
18	81	21.5	270	2 G71061	hypothetical prote
19	81	21.5	396	2 T41405	hypothetical prote
20	81	21.5	409	2 F72504	hypothetical repea
21	80.5	21.4	269	2 T18335	icmG protein - Leg
22	80.5	21.4	647	2 A84265	hypothetical prote
23	80	21.2	102	2 F64341	hypothetical prote
24	80	21.2	103	2 B60608	myosin heavy chain
25	80	21.2	527	2 S33068	myosin heavy chain
26	80	21.2	1225	2 A49464	chromosome segrega
27	80	21.2	1940	2 A59287	myosin heavy chain
28	79.5	21.1	565	1 HMIVE3	hemagglutinin prec
29	79.5	21.1	565	1 HMIVE4	hemagglutinin prec

30	79	21.0	318	2 T49167	hypothetical prote
31	79	21.0	369	1 TVFVAF	transforming prote
32	79	21.0	407	2 F72343	hypothetical prote
33	79	21.0	554	2 G72361	hypothetical prote
34	78.5	20.8	564	1 HMIVE2	hemagglutinin prec
35	78.5	20.8	565	1 HMIVE2	hemagglutinin prec
36	78.5	20.8	1956	2 T16416	hypothetical prote
37	78	20.7	456	2 E86903	hypothetical prote
38	78	20.7	1169	2 A64505	p115 homolog - Met
39	78	20.7	1313	2 A48467	myosin heavy chain
40	78	20.7	1938	1 A40997	myosin heavy chain
41	77.5	20.6	348	2 T04618	heat shock protein
42	77.5	20.6	778	2 T30430	hypothetical prote
43	77.5	20.6	1556	2 F96587	hypothetical prote
44	77	20.4	292	2 I51171	transcription fact
45	77	20.4	338	2 I38567	retrovirus-related

ALIGNMENTS

RESULT 1

G69708

chromosome segregation SMC protein - Bacillus subtilis
N:Alternate names: minichromosome stabilizing protein SMC
C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C/Accession: G69708; JC4819; PC4029

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:9804033; PMID:9384377

A:Accession: G69708

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1186 <KUN>

A:Cross-references: GB:299112; GB:AL009126; NID:g2633902; PIDN:CAB13467.1; PID:g26339

R:Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.

A:Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit,

Gene 172, 17-24, 1996

A:Reference number: JC4819; MUID:96257247; PMID:8654983

A:Accession: JC4819

A:Molecule type: DNA

A:Residues: 1-49,'G',51-161,'G',163-174,'E',176-177,'G',179-191,'G',193-227,'P',229-2
74,'P',476-493,'D',495-514,'D',516-541,'V',543-545,'P',547-585,'SKPLRGSGPAFIISF',601
,F',741-1186 <OGU>

A:Cross-references: DDBJ:D64116; NID:g1389548; PIDN:BAAL0577.1; PID:g1237015

R:Oguro, A.; Kakeshita, H.; Honda, K.; Takamatsu, H.; Nakamura, K.; Yamane, K.

DNA Res. 2, 93-100, 1995

A:Title: Srb: a Bacillus subtilis gene encoding a homologue of the alpha-subunit of t

A:Reference number: JC4093; MUID:96093930; PMID:7584053

A:Accession: PC4029

A:Molecule type: DNA

A:Residues: 1171-1186 <OG2>

A:Cross-references: DDBJ:D49781; NID:g1237017; PIDN:BAA08615.1; PID:g1237018

C:Genetics:

A:Gene: smc

C:Superfamily: chromosome segregation protein SMC1

Query Match

25.1% Score 94.5; DB 2; Length 1186;

C:Species: Gallus gallus (chicken)
C:Date: 20-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 21-Jul-2000
C:Accession: D39371
R:Miller, M.M.; Goto, R.; Young, S.; Chirivella, J.; Hawke, D.; Miyada, C.G.
Proc. Natl. Acad. Sci. U.S.A. 88, 4377-4381, 1991
A:Title: Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex class II genes of Gallus gallus
A:Reference number: A39371; MUID:91239571; PMID:1903541
A:Accession: D39371
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-321 <MIL>
A:Cross-references: GB:M61862; NID:g211259; PIDN:AAA48627.1; PID:g211261

Query Match 22.5%; Score 85; DB 2; Length 321;
Best Local Similarity 26.8%; Pred. No. 4.8;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

Qy 7 HHGHSMAAGDLK-----NKVAQLKRRKVRSLDKAAELKQEVSRLENEIED 52
Db 185 HLAEDLSTADLKLLAAKLVEQREAVERDSQLRKQYEKLGSRATNLTKLLENEIEE 244

Qy 53 LKAKIGDLNNTSGIRRAAKLN 74
Db 245 VEKHLKKI-----GIRAPNKLH 262

RESULT 7
F75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: F75103
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAW>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB50131.1; PID:g545864
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0812
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 22.5%; Score 85; DB 2; Length 880;
Best Local Similarity 37.2%; Pred. No. 13;
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy 18 LKNKVAQLKRRKVRSLDKAAELKQEVSRLENEIEDLKAKIGDL 60
Db 240 LKGIKSELKIQVEKLGKRRKGLKEEKIVQIERSIEKKAKISEL 282

RESULT 8
JQ0894
p115 protein - Mycoplasma hyorhinis
C:Species: Mycoplasma hyorhinis
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 02-Feb-2001
C:Accession: JQ0894
R:Notarnicola, S.M.; McIntosh, M.A.; Wise, K.S.
Gene 97, 77-85, 1991
A:Reference number: JQ0894; MUID:91138990; PMID:1825306
A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding domain of the ATP synthase
A:Accession: JQ0894
A:Molecule type: DNA
A:Residues: 1-979 <NOT>
A:Cross-references: GB:M34956
A:Note: the authors translated the codon AAA for residue 956 as Leu
C:Comment: This protein is located in the cytoplasm.
C:Genetics:
A:Genetic code: SGC3

C:Superfamily: chromosome segregation protein SMC1
C:Keywords: nucleotide binding; P-loop
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 22.5%; Score 85; DB 2; Length 979;
Best Local Similarity 30.9%; Pred. No. 14;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

Qy 17 DLKNKVAQLKRRKVRSL-----DKAAELKQEVSRLENEIEDLKAKIGDLNNTSGI--- 66
Db 323 DQKTIIEIKKQVESLKQINASKQRETELQDLTRLNAKANSKLQENDINKETGVILLE 382

Qy 67 RPPAAKLN 74
Db 383 KKSAAAN 390

RESULT 9
F75216
hypothetical protein PAB2181 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F75216
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: F75216
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49181.1; PID:g5457454
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2181

Query Match 22.4%; Score 84.5; DB 2; Length 281;
Best Local Similarity 40.8%; Pred. No. 4.6;
Matches 20; Conservative 10; Mismatches 12; Indels 7; Gaps 1;

Qy 17 DLKNKVAQLKRRKVRSL-----KVRSLKDKAAELKQEVSRLENEIEDLKAKIG 58
Db 222 ELERKVSLESLNRYETKVKSLKKELEKNKVELSEEVNKLKEGIG 270

RESULT 10
T29145
hypothetical protein F56A3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29145
R:Geisel, C.; Bradshaw, H.; Wamsley, P.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid F56A3.
A:Reference number: Z20578
A:Accession: T29145
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1199 <GEI>
A:Cross-references: EMBL:U80446; PIDN:AAB37802.1; GSPDB:GN00019; CESP:F56A3.4
A:Experimental source: strain Bristol N2; clone F56A3
C:Genetics:
A:Gene: CESP:F56A3.4
A:Map position: 1
A:Introns: 18/1; 80/3; 446/3; 475/1; 607/3; 679/1; 974/2; 1058/1; 1168/3

Query Match 22.3%; Score 84; DB 2; Length 1199;
Best Local Similarity 33.3%; Pred. No. 21;
Matches 17; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

Qy 12 SMASGDLNKKVAQLKRRKVRSLDKAAELKQEVSRLENEIEDLKAKIGDLN 62
Db 198 TLTSGYEERKINDLEAKLLSEIDKVAELEDHIQQRQELDDQSAARLASEN 248

RESULT 11

F90182
hypothetical protein SSO0390 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: F90182
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
arrett, R.A.; Jeffries, A.C.; Kozera, C.-J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: F90182
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <KUR>
A:Cross-references: GB:AEO06641; NID:g13813537; PIDN:AAK40717.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO0390

Query Match	21.8%	Score 82.5;	DB 2;	Length 397;	
Best Local Similarity	34.5%;	Pred. No. 9.4;			
Matches 19;	Conservative	13;	Mismatches 16;	Indels 7;	Gaps 1;

Oy 18 LKNVAOLKKRVSLKDKAELKO-----EVSRLNEIEDLKAKIGDLNNTSG 65
 |.:|.:|.:|.:||| :.|.:|.:|.:|.:|.:|.:|.:|.:|.:|
Db 262 LNNEVSTLRSEISSNSTIASLNKSLANANTQISNLQASEITTLNSEIGKLNVTVG 316

RESULT 12

JC5610
troponin I - sea squirt (Halocyynthia roretzi)
C:Species: Halocyynthia roretzi
C>Date: 23-Sep-1997 #sequence_revision 17-Oct-1997 #text_change 20-Jun-2000.
C:Accession: JC5610
R:Yuasa, H.J.; Sato, S.; Yamamoto, H.; Takagi, T.
J. Biochem. 122, 374-380, 1997
A:Title: Primary structure of troponin I isoforms from the ascidian Halocyynthia roretzi
A:Reference number: JC5610; MUID:98021076; PMID:9378716
A:Accession: JC5610
A:Molecule type: mRNA
A:Residues: 1-173 <YDA>
A:Cross-references: DDBJ:AB001685; NID:g1888344; PIDN:BAA19425.1; PID:g1888345
A:Experimental source: adult Halocyynthia roretzi
C:Comment: This protein binds to actin, and inhibits the interaction between actin and
C:Superfamily: troponin I

Query Match	21.8%	Score 82;	DB 2;	Length 173;	
Best Local Similarity	30.5%;	Pred. No. 4.5;			
Matches 18;	Conservative	16;	Mismatches 23;	Indels 2;	Gaps 1;

Oy 11 GSMAGDCLKNKVAOLKRVRSLDKPAALKEQEVSRLENEIIDLKAKIGDLNNTSGIRRP 69
 |:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
Db 47 GGWSEQDLCLRELHAKIEKVDEQRVDIEVKVNKNQEIEDLNRIFDLRGK--FKRP 103

RESULT 13

A56235
transcription activator MafB - chicken
C:Species: Gallus gallus (chicken)
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 20-Jun-2000
C:Accession: A56235
R: Kataoka, K.; Fujiwara, K.T.; Noda, M.; Nishizawa, M.
Mol. Cell. Biol. 14, 7581-7594, 1994
A:Title: MafB, a new Maf family transcription activator that can associate with Maf and
A:Reference number: A56235; MUID:95021288; PMID:7935473
A:Accession: A56235
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KAT>
A:Cross-references: GB:D28600; NID:g516723; PIDN:BAAA05938.1; PID:g516724

```

C:Genetics:
A:Introns: #status absent
C:Superfamily: maf transforming protein; maf homology
C:Keywords: DNA binding; homodimer; leucine zipper
F:200-289/Domain: maf homology <MAF>

Query Match          21.88; Score 82; DB 2; Length 311;
Best Local Similarity 21.78; Pred. No. 8.1;
Matches 30; Conservative 12; Mismatches 20; Indels 76; Gaps 3;

Qy  5 HHHHHHGSMA$-----G 16
      |||||
Db  159 HHHHHHHQASPTSTSSSSQQLQTSHQHPSSSVEDRFSDQLYSMSVRELNRLRG 218
      ||||| |||||
Qy  17 DLKNVQAQLKKRVSLKDX-----AAELKQEVSRLENE 49
      ||||| |||||
Db  219 FTKDEVIRLKKRRTLKRNQYAQSCRYKRVQOKHLENEKTLQIQVEQLKQEVTRLARE 278
      ||||| |||||
Qy  50 IEDLKAKIGDLNNTSGIR 67
      :|||
Db  279 RDAYLCKEKL-ASNGFR 295
      :|||

RESULT 14
S37894
hypothetical protein YKL072w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL352
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C:Accession: S37894; S37897; S39170; S44515
R:Rasmussen, S.; von Wettstein, D.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37872
A:Accession: S37894
A:Molecule type: DNA
A:Residues: 1-766 <RAS>
A:Cross-references: EMBL:Z28072; NID:g486100; PIDN:CAA81909.1; PID:g486101; MIPS:YKL0
A:Experimental source: strain S288C
R:Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37897
A:Accession: S37897
A:Molecule type: DNA
A:Residues: 1-557 <POH>
A:Cross-references: EMBL:Z28072; MIPS:YKL072w
A:Experimental source: strain S288C
R:Rasmussen, S.W.
submitted to the EMBL Data Library, November 1993
A:Reference number: S39168
A:Accession: S39170
A:Molecule type: DNA
A:Residues: 1-766 <RA2>
A:Cross-references: EMBL:X75780; NID:g433625; PIDN:CAA53402.1; PID:g433628
R:Rasmussen, S.W.
Yeast 10, 69-74, 1994
A:Title: Sequence of a 20.7 kb region of yeast chromosome XI includes the NUP100 gene
rp in addition to seven ORFs with weak or no significant similarity to known proteins
A:Reference number: S44513
A:Accession: S44515
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-766 <RA3>
A:Cross-references: EMBL:X75780; NID:g433625; PIDN:CAA53402.1; PID:g433628
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
C:Genetics:
A:Gene: SGD:STB6
A:Cross-references: SGD:S0001555; MIPS:YKL072w
A:Map position: 11L

```

Query Match 21.8%; Score 82; DB 2; Length 766;
Best Local Similarity 26.7%; Pred. NO. 20;
Matches 20; Conservative 20; Mismatches 25; Indels 10; Gaps 2;

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 1.13584 Seconds
(without alignments)
2702.183 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377

Sequence: 1 MRGSHHHHGHGSMASGLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	94.5	25.1	1186	1 SMC_BACSU	P51834 bacillus su
2	86.5	22.9	323	1 MAFB_RAT	P54842 rattus norv
3	86	22.8	548	1 CEAK_ECOLI	O47502 escherichia
4	85.5	22.7	323	1 MAFB_HUMAN	Q9Y5G3 homo sapien
5	85.5	22.7	323	1 MAFB_MOUSE	P54841 mus musculus
6	85	22.5	880	1 RA50_PYRAB	Q9UZC8 pyrococcus
7	85	22.5	979	1 P115_MYCHR	P41508 mycoplasma
8	82	21.8	766	1 STB6_YEAST	P36085 saccharomyc
9	82	21.8	882	1 RA50_PYRFU	P58301 pyrococcus
10	81	21.5	166	1 YB38_MYCPN	P75260 mycoplasma
11	81	21.5	409	1 PSMR_AERPE	Q9YAC7 aeropyrum p
12	80	21.2	102	1 Y334_METJA	Q57780 methanococc
13	80	21.2	1225	1 SMC1_YEAST	P32908 saccharomyc
14	79.5	21.1	565	1 HEMA_IAHAL	P16994 influenza a
15	79.5	21.1	565	1 HEMA_IAHTO	P17000 influenza a
16	79	21.0	369	1 TMAF_AVISA4	P23091 avian muscu
17	78.5	20.8	564	1 HEMA_IAGRE	P19698 influenza a
18	78.5	20.8	565	1 HEMA_IAHMI	P15658 influenza a
19	78	20.7	442	1 VATC_DROME	Q9V7N5 drosophila
20	78	20.7	1169	1 SMC_METJA	Q59037 methanococc
21	78	20.7	1938	1 MYS_AEQIR	P24733 aequipeten
22	78	20.7	4349	1 DYHC_FUSSO	P78716 fusarium so
23	77	20.4	461	1 US45_LACLC	P22865 lactococcus
24	76.5	20.3	741	1 BSG2_DROME	P11929 drosophila
25	75.5	20.0	550	1 HEMA_IADHK	P43257 influenza a
26	75.5	20.0	1509	1 MYSN_ACACA	P05659 acanthamoeb
27	75	19.9	715	1 CLPB_MYCPN	P75247 mycoplasma
28	75	19.9	1727	1 ALMI_SCHPO	Q9UTK5 schizosacch
29	74.5	19.8	830	1 PAMI_YEAST	P37304 saccharomyc
30	74.5	19.8	879	1 MYSP_DROME	P35415 drosophila
31	74.5	19.8	879	1 RA50_PYRHO	O58687 pyrococcus
32	74	19.6	539	1 MYS3_PYDAT	P39922 hydra atten
33	74	19.6	1957	1 YD86_SCHPO	Q10411 schizosacch

ALIGNMENTS

RESULT 1

SMC_BACSU

ID SMC_BACSU STANDARD; PRT; 1186 AA.

AC P51834; O31735;

DT 01-OCT-1996 (Rel. 34, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chromosome partition protein smc.

GN SMC.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RL Oguero A., Kakeshita H., Takamatsu H., Nakamura K., Yamane K.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Tateuchi M., Tanakashi A., Tanaka T., Terpsstra P., Tognoni K.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RA "The complete genome sequence of the Gram-positive bacterium Bacillus
RA subtilis.";
RA Nature 390:249-256(1997).
RN [3]
RXL Nature 390:249-256(1997) FROM N.A.
RNL SEQUENCE OF 1171-1186 FROM N.A.
RNC STRAIN=168;
RXC MEDLINE=96093930; PubMed=7584053;

34 74 19.6 2022 1 ANTI_ONCVO P21249 onchocerca
35 73.5 19.5 550 1 HEMA_IABAN P03441 influenza a
36 73.5 19.5 550 1 HEMA_IADH1 P12582 influenza a
37 73.5 19.5 550 1 HEMA_IADH2 P12583 influenza a
38 73.5 19.5 550 1 HEMA_IADH3 P12584 influenza a
39 73.5 19.5 550 1 HEMA_IADH4 P12585 influenza a
40 73.5 19.5 550 1 HEMA_IADH6 P12587 influenza a
41 73.5 19.5 550 1 HEMA_IADH7 P12588 influenza a
42 73.5 19.5 550 1 HEMA_IADHL P43258 influenza a
43 73.5 19.5 550 1 HEMA_IADHM P43259 influenza a
44 73.5 19.5 550 1 HEMA_IAME6 P12589 influenza a
45 73.5 19.5 550 1 HEMA_IADH3 P11134 influenza a

RA Oguro A., Kakeshita H., Honda K., Takamatsu H., Nakamura K.,
 RA Yamane K.;
 RT "srb: a Bacillus subtilis gene encoding a homologue of the alpha-
 RT subunit of the mammalian signal recognition particle receptor.";
 RL DNA Res. 2:95-100(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=98367134; PubMed=9701812;
 RA Moriya S., Tsujikawa E., Hassan A.K., Asai K., Kodama T.,
 RA Ogawara N.;
 RT "A Bacillus subtilis gene-encoding protein homologous to eukaryotic
 RT SMC motor protein is necessary for chromosome partition.";
 RL Mol. Microbiol. 29:179-187(1998).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=98241483; PubMed=9573042;
 RA Britton R.A., Lin D.C., Grossman A.D.;
 RA "Characterization of a prokaryotic SMC protein involved in chromosome
 RT partitioning.";
 RL Genes Dev. 12:1254-1259(1998).
 CC -|- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
 CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION.
 CC -|- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D64116; BAA10977.1; -;
 DR EMBL; Z99112; CAB13467.1; -;
 DR EMBL; D49781; BAA08615.1; -;
 DR Subtilist; BG11538; smc
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam; PF02463; SMC_N; 1.
 DR Pfam; PF02483; SMC_C; 1.
 KW ATP-binding; Coiled coil; Complete proteome.
 FT NP_BIND 31 38
 FT DOMAIN 244 481
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 664 943
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 989 1031
 FT COILED COIL (POTENTIAL).
 FT E -> G (IN REF. 1).
 FT E -> G (IN REF. 1).
 FT K -> E (IN REF. 1).
 FT E -> G (IN REF. 1).
 FT E -> G (IN REF. 1).
 FT A -> P (IN REF. 1).
 FT K -> E (IN REF. 1).
 FT A -> P (IN REF. 1).
 FT A -> P (IN REF. 1).
 FT D -> G (IN REF. 1).
 FT N -> D (IN REF. 1).
 FT D -> D (IN REF. 1).
 FT KEELSKQ -> TRRAFEA (IN REF. 1).
 FT E -> K (IN REF. 1).
 FT E -> K (IN REF. 1).
 FT I -> F (IN REF. 1).
 FT A -> P (IN REF. 1).
 FT E -> D (IN REF. 1).
 FT E -> D (IN REF. 1).
 FT L -> V (IN REF. 1).
 FT A -> P (IN REF. 1).
 FT QSRDAETAAHSSFL -> SKPLRNGSGPAFIISF (IN
 REF. 1).
 FT TVLTIEDLK -> NRSDYRLKLG (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT S -> T (IN REF. 1).
 FT E -> G (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT CONFLICT 623
 FT CONFLICT 664
 FT CONFLICT 676
 FT CONFLICT 680
 FT CONFLICT 694

FT CONFLICT 701 701 K -> Q (IN REF. 1).
 FT CONFLICT 726 726 L -> V (IN REF. 1).
 FT CONFLICT 738 740 LQV -> POF (IN REF. 1).
 SQ SEQUENCE 1186 AA; 135510 MW; 0163227A2BCA2B CRC64;
 Query Match 25.1%; Score 94.5; DB 1; Length 1186;
 Best Local Similarity 36.8%; Pred. No. 0.79;
 Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;
 QY 11 GSNASGDKKNKVAQLKRVSLKD-----KAEELKQEVSRLENETEDLKAKIGDL 60
 DB 658 GSGTGAIVKKNNSLLGRSELEDVTKRLAEEMKTALEQEVTKLHKSQDMKKLADL 717
 QY 61 NNT-SGIR 67
 DB 718 RETGEGLR 725
 RESULT 2
 MAFB_RAT
 ID MAFB_RAT STANDARD; PRT; 323 AA.
 AC P54842;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma
 DE oncogene homolog B) (Transcription factor MAF1).
 GN MAFB OR MAF1
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Liver;
 RX MEDLINE=97190228; PubMed=9038383;
 RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
 RA Kuboki Y., Nishizawa M., Nishi S.;
 RT "Rat maf related genes: specific expression in chondrocytes, lens and
 RT spinal cord.";
 RL Oncogene 14:745-750(1997).
 CC -|- FUNCTION: Plays a pivotal role in regulating lineage-specific
 CC hematopoiesis by repressing Ets1-mediated transcription of
 CC erythroid-specific genes in myeloid cells (By similarity).
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U56241; AAB50062.1; -;
 DR InterPro: IPR004826; TF_Maf.
 DR InterPro: IPR004827; TF_BZIP.
 DR Pfam; PF03131; bzip_Maf; 1.
 DR SMART; SM00338; BRLZ; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 FT DNA_BIND 238 264
 FT BASIC MOTIF.
 FT DOMAIN 266 287
 FT LEUCINE-ZIPPER.
 FT DOMAIN 131 143
 FT POLY-HIS.
 FT DOMAIN 158 167
 FT POLY-HIS.
 FT DOMAIN 194 201
 FT POLY-ALA.
 SQ SEQUENCE 323 AA; 35792 MW; 6E386340D1F840A5 CRC64;
 Query Match 22.9%; Score 86.5; DB 1; Length 323;
 Best Local Similarity 23.1%; Pred. No. 1;
 Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;
 QY 5 HHHHHH-----GSMAS----- 15


```

DR InterPro: IPR004826; TE_Maf.
DR Pfam: PF03131; bZIP_Maf; 1.
DR SMART: SM00338; BRLZ; 1.
DR KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA_BIND 238 264 BASIC MOTIF.
FT DOMAIN 266 287 LEUCINE-ZIPPER.
FT DOMAIN 131 143 POLY-HIS.
FT DOMAIN 158 167 POLY-HIS.
FT CONFLICT 52 52 A -> V (IN REF. 1).
FT CONFLICT 241 241 Q -> H (IN REF. 1).
SQ SEQUENCE 323 AA; 35792 MW; A0F3C09FB936CB16 CRC64;

Query Match 22.7%; Score 85.5; DB 1; Length 323;
Best Local Similarity 23.1%; Pred. No. 1,2;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

QY 5 HHHHHH-----GSMAS----- 15
||||| |||
DB 162 HHHHHHQQASPPSSAASPAQQLPTSHQPGPHATASATAAGNGSVEDRFDDQLVMSV 221
QY 16 -----GDLKNVAQLKKRVSLDK-----AALK 40
| : : | : : | : : | : :
DB 222 RELNHLRLGFTKDEVIRLQKRRRLTKNGRYAQSCKYKRVQKHHLNEKTQLIQQVEQLK 281
QY 41 QEVSRLENEIDLKAKIGDLNLTSGIR 67
||||| | : | : | : | : |
DB 282 QEVSRLARERDAYKVEKELAN-SGFR 307

RESULT 5
MAFB_MOUSE
ID MAFB_MOUSE STANDARD; PRT; 323 AA.
AC P54841;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma
DE oncogene homolog B) (transcription factor MAF1) (segmentation protein
DE KR) (Kreisler).
DE MAFB OR MAF1 OR KRML.
DE MS Mus musculus (Mouse).
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DE NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RN MEDLINE-95094266; PubMed-8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1023-1034(1994).
CC -|- FUNCTION: Plays a pivotal role in regulating lineage-specific
CC hematopoiesis by repressing Ets1-mediated transcription of
CC erythroid-specific genes in myeloid cells.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.
CC -|- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND
CC IN THE CAUDAL HINDBRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),
CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT
CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY
CC THROUGH R6.
CC -|- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; L36435; AAA65689.1; --
CC TRANSFAC; T01439; --

```

```

DR SMART: SM00382; AA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 144 745 COILED COIL (POTENTIAL).
SQ SEQUENCE 880 AA; 103970 MW; FDBI17EC7E026479 CRC64;

Query Match 22.5%; Score 85; DB 1; Length 880;
Best Local Similarity 37.2%; Pred. No. 3.7;
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 18 LKNKVAOLKKRVSRSLDKAAELKQEVSRLENEIEDLKAKIGDL 60
    | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 240 IKGKISELKQVEKLGKRRKKGLEKIVQIERISIEEKKAKISEL 282

RESULT 7
P115_MYCHR STANDARD; PRT; 979 AA.
AC F41508;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P115 protein.
OS Mycoplasma hyorhinitis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2109;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=91138950; PubMed=1825306;
Notarnicola S.M., McIntosh M.A., Wise K.S.;
"A Mycoplasma hyorhinitis protein with sequence similarities to
nucleotide-binding enzymes.";
Gene 97:77-85(1991).
RL CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC -1- FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC P115.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
-----
ENBL: M34956; AAA25423.1; -.
PIR: JQ0894; JQ0894.
InterPro: IPR003439; ABC_transportr.
InterPro: IPR005289; GTP-binding_dom.
InterPro: IPR003405; SMC_C.
InterPro: IPR003395; SMC_N.
Pfam: PF02463; SMC_N; 1.
Pfam: PF02483; SMC_C; 1.
TIGRFS: TIGR00650; MG442; 2.
ATP-binding: Coiled coil.
NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 169 224 COILED COIL (POTENTIAL).
FT DOMAIN 231 400 COILED COIL (POTENTIAL).
FT DOMAIN 569 821 COILED COIL (POTENTIAL).
FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
SQ SEQUENCE 979 AA; 110566 MW; 30D51C56B56280F4 CRC64;

Query Match 22.5%; Score 85; DB 1; Length 979;
Best Local Similarity 30.9%; Pred. No. 4.3;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

QY 17 DLKNKVAOLKKRVSRSL-----KDKAAELKQEVSRLENEIEDLKAKIGDLNNTSGI--- 66
    | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 323 DQYKTEIEIKKQVESLKIQINASKQREIEDQQUTRNANKANSUKQENDINKEIGVLE 382

QY 67 RRPAAKLN 74

```

```
ID RA50_PYRFU STANDARD; PRT; 882 AA.
AC P58301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PF1167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RN SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC STRAIN-Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
RA Carney J.P.;
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical
RT characterization reveal an evolutionarily conserved multiprotein
RT machine.";
RL J. Bacteriol. 182:6036-6041(2000).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RT Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN [3]
RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC STRAIN-Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20348838; PubMed=10892749;
RA Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA Carney J.P., Tainer J.A.;
RT "Structural biology of Rad50 ATPase: ATP-driven conformational
RT control in DNA double-strand break repair and the ABC-ATPase
RT superfamily.";
RL Cell 101:789-800(2000).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC Rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site.
CC -1- SUBUNIT: Forms a complex with mrell.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF010225; AAL81291.1; -
DR PDB; 1F2T; 02-AUG-00.
DR PDB; 1F2U; 02-AUG-00.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF02463; SMC_N.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
KW Complete proteome.
FT NP_BIND 30 37 ATP.
FT DOMAIN 148 744 COILED COIL (POTENTIAL).
FT SEQUENCE 882 AA; 103839 MW; 3ADCBD250382A99E CRC64;
Query Match 21.8%; Score 82; DB 1; Length 882;
Best Local Similarity 29.1%; Pred. No. 6.7;
Matches 16; Conservative 15; Mismatches 24; Indels 0; Gaps 0;
QY 11 GSMASGDLKNKVAQLKRRKSLDKAAELKQVSRLENEIEDLKAKIGDINNTSG 65
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 388 GDSKSPDIKKLLELETKTTIEERNEITQIGELKNKIGDLKTAIELKRAKG 442
```

```
RESULT 10
YB38_MYCPN STANDARD; PRT; 166 AA.
AC P75260;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN138 (E07_orf166).
GN MPN138 OR MP016.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0134 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000003; AAB95664.1; -
DR InterPro; IPR002862; DUF16.
DR Pfam; PF01519; DUF16; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 166 AA; 19518 MW; BE4F5377B2FA709 CRC64;
Query Match 21.5%; Score 81; DB 1; Length 166;
Best Local Similarity 37.2%; Pred. No. 1.5;
Matches 16; Conservative 15; Mismatches 12; Indels 0; Gaps 0;
QY 18 LKNKVAQLKRRKSLDKAAELKQVSRLENEIEDLKAKIGDL 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 81 IENKVDKLEKVKDKLEKVKLEAKVDKLEKVKLEAKVDKL 123
RESULT 11
PSMR_AERPE STANDARD; PRT; 409 AA.
AC Q9YAC7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN PAN OR APE2012.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
```

```
CC CC -1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC CC proteins, but not small peptides, by the 20S proteasome (By
CC CC similarity).
CC CC -1- SUBUNIT: Homohexameric (Potential).
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; AP000063; BAAB1022.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; FALSE_NEG.
DR TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.58; Score 81; DB 1; Length 409;
Best Local Similarity 37.38; Pred. No. 3.7;
Matches 19; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 3 GSHHHHHGMSAGDLKNKVAQLKRVSLKDKAAELKQVSRLENEIDL 53
DB 8 GSRSHRHNGGHSRDEVEIRI--LADKVRSLTKKELISLQKELEYKNEITKL 56

RESULT 12
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M70334.
GN M70334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U67487; AAB98322.1; -.

-1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
proteins, but not small peptides, by the 20S proteasome (By
similarity).
-1- SUBUNIT: Homohexameric (Potential).
-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP000063; BAAB1022.1; -.
InterPro; IPR003593; AAA_ATPase.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
PROSITE; PS00674; AAA; FALSE_NEG.
TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.58; Score 81; DB 1; Length 409;
Best Local Similarity 37.38; Pred. No. 3.7;
Matches 19; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 3 GSHHHHHGMSAGDLKNKVAQLKRVSLKDKAAELKQVSRLENEIDL 53
DB 8 GSRSHRHNGGHSRDEVEIRI--LADKVRSLTKKELISLQKELEYKNEITKL 56

RESULT 12
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M70334.
GN M70334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U67487; AAB98322.1; -.

-1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
proteins, but not small peptides, by the 20S proteasome (By
similarity).
-1- SUBUNIT: Homohexameric (Potential).
-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP000063; BAAB1022.1; -.
InterPro; IPR003593; AAA_ATPase.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
PROSITE; PS00674; AAA; FALSE_NEG.
TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.28; Score 80; DB 1; Length 102;
Best Local Similarity 40.08; Pred. No. 1.1;
Matches 22; Conservative 8; Mismatches 17; Indels 8; Gaps 1;

QY 19 KKKVAQLKRVSR-----SLKDKAAELKQVSRLENEIDLKAKIGDLNNTSG 65
DB 37 KKKVADLVKRVMEQFHEVSNQEKIAELKAEELKEELKEIRGLKAGKAG 91

RESULT 13
SMC1_YEAST
ID SMC1_YEAST STANDARD; PRT; 1225 AA.
AC P32908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YEL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Strunnikov A.V., Larionov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family.";
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; L00602; AAA16595.1; -.
DR EMBL; D50617; BAA09230.1; -.
DR PIR; A49464; A49464.
DR PIR; S41804; S41804.
DR SGD; S0001886; SMC1.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Mitosis; ATP-binding; Coiled coil; Nuclear protein.
```

```
FT NP_BIND 33 40 ATP (POTENTIAL).
FT DOMAIN 173 489 COILED COIL (POTENTIAL).
FT DOMAIN 679 1063 COILED COIL (POTENTIAL).
FT DOMAIN 1057 1061 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1137 1164 ALA/ASP-RICH (DA-BOX).
FT MUTAGEN 173 173 S->L: IN TS MUTANT SMCI-2.
FT MUTAGEN 458 458 N->D: IN TS MUTANT SMCI-1.
SQ SEQUENCE 1225 AA; 141279 MW; B504017AA0ECC8C CRC64;

Query Match 21.1%; Score 80; DB 1; Length 1225;
Best Local Similarity 33.9%; Pred. No. 14;
Matches 19; Conservative 12; Mismatches 21; Indels 4; Gaps 1;

QY 12 SMASGDLNKKVAQLKRRVSLKDKAAHLKQEVSRLENEIEB----DLKAKIGDGLNNT 63
Db 718 SLNLSIDIANLRVTQOKRSLDENRLEIKYVHNDLIEIKPQKTELKKLDDLENT 773

RESULT 14
HEMA_IHAHAL STANDARD; PRT; 565 AA.
AC P16994; Q83990;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Algiers/72).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawasaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24721; AAA43100.1; ALT_SEQ.
CC PIR; D34064; HMIVB4.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 340 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 33.9%; Pred. No. 14;
Matches 19; Conservative 12; Mismatches 21; Indels 4; Gaps 1;
```

```
Best Local Similarity 28.1%; Pred. No. 7;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

QY 3 GSHHHHHGSMASGDLNKKVA-----QLKRRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 367 GFRHQNSEGTGQAGDLKSTQAAIDQINGKLNRVIEKTNKFKHQIEKFESEVEGRIOQLEK 426
QY 56 KIGD 59
Db 427 YVED 430

RESULT 15
HEMA_IHAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawasaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24720; AAA43111.1; ALT_SEQ.
CC PIR; C34064; HMIVE3.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 565 AA; 63580 MW; 84B7D4AD70629B7A CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 7;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

QY 3 GSHHHHHGSMASGDLNKKVA-----QLKRRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 367 GFRHQNSEGTGQAGDLKSTQAAIDQINGKLNRVIEKTNKFKHQIEKFESEVEGRIOQLEK 426
QY 56 KIGD 59
Db 427 YVED 430

RESULT 15
HEMA_IHAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawasaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24720; AAA43111.1; ALT_SEQ.
CC PIR; C34064; HMIVE3.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 565 AA; 63580 MW; 84B7D4AD70629B7A CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 7;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

QY 3 GSHHHHHGSMASGDLNKKVA-----QLKRRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 367 GFRHQNSEGTGQAGDLKSTQAAIDQINGKLNRVIEKTNKFKHQIEKFESEVEGRIOQLEK 426
QY 56 KIGD 59
Db 427 YVED 430
```

Db 367 GFRHONSEGTGAGDLKSTQAAIDQINGKLN RVIEKTNEKFHQIEKEFSEVEGRIQDLEK 426

QY 56 KIGD 59

: |

Db 427 YVED 430

Search completed: December 18, 2002, 16:18:25
Job time : 3.13584 secs

This Page Blank (uspto)


```

AC O8TV66;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Predicted secreted protein.
GN MK1532.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva V.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Shcherbinina O.V., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Natarle D.A., Koonin E.V., Kozayavkin S.A.;
RA Malykh A.G.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010443; AAM02745.1; -.
KW Complete proteome.
SQ SEQUENCE 615 AA; 66941 MW; 77729702806BC922 CRC64;

Query Match 24.1%; Score 91; DB 17; Length 615;
Best Local Similarity 38.8%; Pred. No. 6;
Matches 19; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

QY 15 SGDLKKNKVAQLKPKVSLKDKAELKQEVSRLENEIEDLKAKIGDLNNT 63
Db 528 SNELTQTTQLQNKVSTVLEQGGELKQEVSKTSQEVSLKGSIDECKNT 576

RESULT 3
Q98UK5 PRELIMINARY; PRT; 356 AA.
ID Q98UK5
AC Q98UK5;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Transcription factor MafB.
GN MAFB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064923; PubMed=11134968;
RA Kajihara M., Kawauchi S., Kobayashi M., Ogino H., Takahashi S.,
RA Yasuda K.;
RT "Isolation, Characterization, and Expression Analysis of Zebrafish
RT Large Mafs."
RL J. Biochem. 129:139-146(2001).
DR EMBL; AB006322; BAB21102.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40233 MW; DE4C96B62C058865 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
||||| |
Db 215 HHHHHHPHGGQGHGCGGLNVEDRFSDDQLVTMSVRELNRHLRGFTKDEVIRLKQK 274
||||| |
QY 29 VRSLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
||||| |
Db 275 RRTLKNRGYAQSCRFRVQVKHLENEKTQLINQVEQLKQEIINRLARERDAYKLKCEKLT 334
||||| |
QY 62 NTSGIR 67
Db 335 GANGFR 340

RESULT 5
Q98UK3 PRELIMINARY; PRT; 315 AA.
ID Q98UK3
AC Q98UK3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Smafl.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064923; PubMed=11134968;
RA Kajihara M., Kawauchi S., Kobayashi M., Ogino H., Takahashi S.,
RA Yasuda K.;
RT "Isolation, Characterization, and Expression Analysis of Zebrafish
RT Large Mafs."
RL J. Biochem. 129:139-146(2001).
DR EMBL; AB006322; BAB21102.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40233 MW; DE4C96B62C058865 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
||||| |
Db 215 HHHHHHPHGGQGHGCGGLNVEDRFSDDQLVTMSVRELNRHLRGFTKDEVIRLKQK 274
||||| |
QY 29 VRSLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
||||| |
Db 275 RRTLKNRGYAQSCRFRVQVKHLENEKTQLINQVEQLKQEIINRLARERDAYKLKCEKLT 334
||||| |
QY 62 NTSGIR 67
Db 335 GANGFR 340

```

```

Db 275 RRTLKNRGYAQSCRFRVQVKHLENEKTQLINQVEQLKQEIINRLARERDAYKLKCEKLT 334
QY 62 NTSGIR 67
Db 335 GANGFR 340

RESULT 4
Q98UK3 PRELIMINARY; PRT; 356 AA.
ID Q98UK3
AC Q98UK3;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Transcription factor Val.
GN VAL OR VALENTINO.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165393; PubMed=9425134;
RA Moens C.B., Cordes S.P., Giorgianni M.W., Barsh G.S., Kimmel C.B.;
RT "Equivalence in the genetic control of hindbrain segmentation in fish
RT and mouse."
RL Development 125:381-391(1998).
DR EMBL; AF006641; AAC18821.1; -.
DR ZFIN; ZDB-GENE-980526-515; val.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40243 MW; 07420DB0F6CD08F1 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
||||| |
Db 215 HHHHHHPHGGQGHGCGGLNVEDRFSDDQLVTMSVRELNRHLRGFTKDEVIRLKQK 274
||||| |
QY 29 VRSLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
||||| |
Db 275 RRTLKNRGYAQSCRFRVQVKHLENEKTQLINQVEQLKQEIINRLARERDAYKLKCEKLT 334
||||| |
QY 62 NTSGIR 67
Db 335 GANGFR 340

RESULT 5
Q98UK3 PRELIMINARY; PRT; 315 AA.
ID Q98UK3
AC Q98UK3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Smafl.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064923; PubMed=11134968;
RA Kajihara M., Kawauchi S., Kobayashi M., Ogino H., Takahashi S.,
RA Yasuda K.;
RT "Isolation, Characterization, and Expression Analysis of Zebrafish
RT Large Mafs."
RL J. Biochem. 129:139-146(2001).
DR EMBL; AB006322; BAB21102.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40233 MW; DE4C96B62C058865 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
||||| |
Db 215 HHHHHHPHGGQGHGCGGLNVEDRFSDDQLVTMSVRELNRHLRGFTKDEVIRLKQK 274
||||| |
QY 29 VRSLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
||||| |
Db 275 RRTLKNRGYAQSCRFRVQVKHLENEKTQLINQVEQLKQEIINRLARERDAYKLKCEKLT 334
||||| |
QY 62 NTSGIR 67
Db 335 GANGFR 340

```

```

J. Biochem. 129:139-146(2001)).
DR EMBL; AB006324; BAB21104.2; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bzip_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 315 AA; 35705 MW; E0147093BAC67CE2 CRC64;

Query Match                23.6%; Score 89; DB 13; Length 315;
Best Local Similarity      25.5%; Pred.No. 4;6;
Matches 24; Conservative 12; Mismatches 20; Indels 38; Gaps

Qy 5 HHHHHGSMASGD-----KKNVAQLKKRVSLDKAAE- 38
   ||||| | |
Db 179 HHHHHGHHAHARLRFSDQVLVMTVRELNQLRGFSKEEVIRLKQRRTLKNGRYAQ 238
   ||||| | |

Qy 39 -----LKQEVSRLENIEDLKAKIGDL 60
   | | | : | : | : | : |
Db 239 SCYKRVRQQHMLESEKCTLSQVEQLKDQVARL 272
   | | | : | : | : | : |

RESULT 6
QI91891 PRELIMINARY; PRT: 146 AA.
AC QI91891
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE bZIP transcription factor L-MAF (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX Gallus.
OQ NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-BRED FAYOUMI, AND BREED G-B1 LEHORN; TISSUE-SPLEEN;
RA Zhou H., Lamont S.J.;
RT "Genetic variation among chicken lines and mammalian species in
RT specific genes";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBDJ databases.
RM EMBL; AF221558; AAF71240.1; -.
DR EMBL; AF221557; AAF71239.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bzip_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
FT NON_TER 1 1
FT NON_TER 146 146
FT FT
SQ SEQUENCE 146 AA; 17579 MW; 382BAD1D098E1DCC CRC64;

Query Match                23.5%; Score 88.5; DB 13; Length 146;
Best Local Similarity      27.5%; Pred.No. 2;3;
Matches 28; Conservative 11; Mismatches 14; Indels 49; Gaps

Qy 4 SHHHHHHG-----SNASGD-----KKNVAQLKKRVSLDK----- 35
   ||||| | | | : |
Db 39 AHHHHHHHHLERPFSDQLVMSVRELNQLRGFSKEEVIRLKQRRTLKNGRYAQSC 98
   ||||| | | | : |

Qy 36 -----AELKQEVSRLENIEDLKAK 56
   ||||| | | | : |
Db 99 RYKRVRQHILENEKCOLSQVEQLKQEVSRLENERDYLYEK 140
   ||||| | | | : |

RESULT 7
O57342 PRELIMINARY; PRT: 286 AA.
ID O57342
AC O57342;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DI 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE bZIP transcription factor MafA.
GN MAFA.
```


THIS PAGE BLANK (USPTO)